Using Real-time Data to Improve Reliability on High-Frequency Transit Services

by

David Maltzan

B.A., Mathematics and Economics
Tufts University (2009)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author . . . .
Department of Civil and Environmental Engineering
May 21, 2015

Certified by . . . .
Nigel H.M. Wilson
Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

Certified by . . . .
John P. Attanucci
Research Associate, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by . . . .
Heidi M. Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Departmental Committee for Graduate Students

Using Real-time Data to Improve Reliability on High-Frequency Transit Services

by

David Maltzan

Submitted to the Department of Civil and Environmental Engineering
on May 21, 2015, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

In recent years, automatically-collected data from many transit agencies have been made available to the public in real time. This has dramatically improved the experience of riding transit, by allowing passengers to use detailed information on the current state of service to make more informed travel decisions. The "open data" movement has allowed independent mobile-phone app developers to create a variety of useful tools to improve the passenger experience. However, agencies' use of real-time data for operational purposes has lagged behind customer-facing app development.

This research examines the use of real-time data for the application of operational control strategies on transit services. Two high-frequency bus routes of the Massachusetts Bay Transportation Authority are used as a case study. It begins with the development of an application to download, interpret, and present data on bus service and recommended control actions in a graphical user interface. This application is then used to conduct an experiment with a terminal-based holding strategy on MBTA Route 1. The results of this experiment drive further investigation into the causes of deviations from scheduled or assigned departure times at terminals. To supplement the experimental data, a simulation model of MBTA Routes 1 and 28 is developed. This simulation is used to test additional control strategies, as well as the effect of reducing unexplained operator deviations from assigned departure times.

The research finds that real-time data can be used to create significant operational improvements. In particular, holding strategies at terminals, along with reducing unexplained operator deviations from assigned terminal departure times, have a strong effect. Several specific recommendations are made for a number of strategies that the MBTA can use to improve the precision of terminal departure times on bus services. This research also finds that holding at midpoints and short-turning can provide some additional benefit, but the costs and benefits to passengers of these strategies are more complicated and should be investigated with further research and implemented using optimization schemes rather than the heuristic rules used here.

Thesis Supervisor: Nigel H.M. Wilson
Title: Professor, Department of Civil and Environmental Engineering

Thesis Supervisor: John P. Attanucci
Title: Research Associate, Department of Civil and Environmental Engineering

Acknowledgments

Thanks are due first to my advisors, Nigel Wilson, John Attanucci, and David Block-Schachter, who guided this work from start to finish, and from whom I have learned far more than any coursework could teach me. Their high standards and attention to detail were invaluable, and they helped me achieve more than I would have thought possible.

Thanks to Haris Koutsopoulos, Fred Salvucci, Mikel Murga, Jinhua Zhao, and Jay Gordon, whose insightful comments during lab meetings contributed to this work. A huge thanks to Gabriel Sanchez-Martinez, not only for providing well-formatted, well-documented code for me to modify willy-nilly, but for taking time out of his own work for many, many, detailed white-board discussions on transit control strategies.

To Dave Carney and the inspectors, operators, and dispatchers of MBTA Bus Operations who cooperated with my experiment, answered many questions, and provided valuable insight into the on-the-ground realities of running an urban bus service.

To Dominick Tribone, Dave Barker, Laurel Paget-Seekins, and many other MBTA staff members who provided thoughtful comments and ideas for research, and hosted me over the summer at 10 Park Plaza.

To my fellow students on the MBTA project: Mike, Catherine, Raph, William, Katie, and Chris. Thanks for working together to make the most out of our massive datasets, to get all our PowerPoint presentations coordinated, and to make sure we all showed up at 45 High Street on time.

To Emily, Lauren, Dan, Nat, Zach, Anne, Becca, and the rest of the M.S.T. class of 2015. I couldn't have asked for a better group to go on this journey with. Thanks for being there for all the fun times, and all the tough times.

To MIT Medical's Mental Health and Counseling service - they don't get thanked enough, but they provide amazing support for students and staff in a place where it is all-too-frequently needed.

To Michael and Michelle, for being such kind and understanding roommates, even as I came and went at all hours, forgot to wash the dishes, and committed other roommate sins.

To Nick, for being my gym buddy and listening to my complaints, even when you had your own (far more stressful) problems to deal with. To Ross, for listening to me from near and far, and lending your hitherto-unknown editing talents.

To the Garimella family, for giving me a second home. Sometimes a home-cooked dinner can be a lifesaver.

To my Dad, who rode the 62 to Alewife every day. To my brother Pete, who inspires me with the work and commitment he puts in at his start-up every day. To my Mum, without whom I would neither have gotten into, nor out of, MIT. Thanks for always checking in on me.

Finally, to Swini. Thanks for listening to me when I needed to talk, and letting me be silent when I needed not to talk. Thanks for organizing our trips to the BSO, and other little things that gave me something to look forward to along the way. You're the best.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This research examines the application of control strategies for improving reliability on high-frequency bus routes using real-time Automated Vehicle Location (AVL) data. Improving reliability has several components, including operations planning, control strategies, and staff behavior and attitudes. Performance data from both regular service and an experiment, as well as simulations, are used to determine a feasible and effective strategy for controlling buses to improve reliability.

1.1 Motivation

This research is motivated by the importance of reliability to the passenger experience on high-frequency bus services, as well as the availability of technology that allows easy access to real-time bus location data.

1.1.1 The importance of reliability

Reliability on high-frequency bus routes is an ongoing problem for transit agencies. Factors such as traffic, weather, passenger arrival rates at stops, and operator behavior are all highly variable and difficult to predict, leading to a high variability in travel times. Variability in travel time in turn leads to a pattern commonly known as "bunching", in which a bus that has been delayed encounters further delays due to the increased build-up of passengers at stops ahead of it, while its following bus speeds up due to picking up fewer passengers, until the two buses are bunched together (Newell and Potts, 1964). This leads to poor service for passengers who face longer wait times than if headways were even, and are more likely to experience very crowded buses. It even leads to poor perception of service among those who have a short wait, but observe multiple buses arriving together.

1.1.2 Technological progress

Control strategies for dealing with bus bunching have been studied through models, simulations, and experiments for many years. However, it is only recently that data on locations of buses has become generally available in real time. Using real-time data, researchers have tested different control strategies, often applied by supervisors receiving radio instructions from dispatchers at a control center (Pangilinan et al, 2008, Strathman et al, 2001, and Bartholdi and Eisenstein, 2012). Even more recently, the availability of cheap, easy-to-use mobile devices with fast Internet access has opened up even more options for implementing control strategies.

1.2 Objectives

The goal of this research is to develop and test control strategies which are both effective and easily implemented given the constraints faced by typical American transit agencies, and the technology available today. These constraints include limited supervisory personnel with heavy workloads, crowded bus terminals with many routes intersecting, long and variable bus dwell times due to variability in boarding passengers, and a low tolerance of passengers for disruptions to their expected service. This thesis will examine both the effectiveness of particular strategies, and also the operational challenges that must be dealt with to control bus operations more effectively.

1.3 Approach

This section summarizes the approach of this thesis, which involves analysis of AVL data, experimental testing, and simulation modeling.

1.3.1 Automated tool development

The approach begins with the development of an automated decision tool. We first describe four types of control strategies (holding, deadheading, expressing, and short-turning), and the data requirements for implementation of the various strategies. We then develop a design for a generic software application to use real-time data to implement a control strategy on a transit service, including the necessary components and the options that are available for each. Finally, we present a specific application developed for a specific case: an experimental test of a holding strategy on MBTA Route 1.

1.3.2 Experimental approach

Next, we conduct an experiment using the automated tool to control buses on Route 1. The results of the experiment shed light on the effectiveness of the strategy, and also give important insight into the challenges faced when implementing control strategies. We also gather supporting data on the environment at the terminals, including scheduled and actual cycle times, operator behavior on breaks, and passenger loading times, to better understand these challenges. Some of the problems identified, such as issues with the terminal layout, are specific to Route 1. However, many of the issues identified are common to many routes, and the analytical approach applied can be used to find similar issues on other routes.

1.3.3 Simulation modeling

Finally, we use simulation to test more possible approaches, varying the route, the type of control (adding midpoint-holding and short-turn options), and operator behavior (varying the level of adherence to assigned departure times by operators). The simulation approach complements the experimental approach, providing data to support potential future experiments. The midpoint-holding and short-turn strategies applied are specific to the individual routes, but similar strategies can be found on many bus routes in the MBTA system and elsewhere.

1.3.4 MBTA application context

This thesis will use case studies and examples from the Massachusetts Bay Transportation Authority (MBTA). The MBTA is the primary public transportation operator in the Greater Boston region, operating bus, rapid transit, commuter rail, and ferry services. MBTA Bus Operations operates 174 bus routes, including local, express, and bus rapid transit (BRT) routes, and covering areas ranging from the inner city to distant suburbs. Fifteen of these are designated as "Key Bus Routes", which are local routes that serve high-density corridors with heavy demand for service on all days of the week (MBTA Service Delivery Policy, 2010). The Key Routes are operated with longer spans of service and at higher frequencies than other local routes. In this thesis, two Key Routes in particular will be examined: Route 1, a crosstown route running from Cambridge to Roxbury, and Route 28, a radial route running from Mattapan to Ruggles Station.

1.4 Outline of Thesis

Chapter 2 reviews the literature on service reliability and bus supervision, models of bus service control, and experiments with controls. This literature informs and sets the stage for the research done for this thesis.

Chapter 3 covers the use of automated tools for transit performance management. Various control strategies are described, along with the data requirements for implementing them. A generic design for a software application to implement a control strategy is described, followed by a specific implementation of this design in the context of MBTA Route 1.

Chapter 4 describes an experiment performed on the MBTA's Route 1 bus route. A basic description of Route 1 is given, with baseline reliability conditions established using AVL data. An initial test run as well as the weeklong experiment are described using results from the AVL data as well as qualitative observations of operator and supervisor behavior. Following the experiment results, the chapter examines obstacles to the implementation of terminal control strategies. Behavior of bus operators and supervisors, the layout of Dudley Station, and the boarding time of passengers are all discussed as possible causes of ineffective dispatching at Dudley.

In Chapter 5, a simulation of bus routes is developed, based on simulation models by Sanchez-Martinez (2014) and Milkovits (2008). This simulation is validated against two MBTA bus routes. In Chapter 6, the simulation is used to test various additional strategies that could be implemented using real-time data, including holding at midpoints and short-turning. It is also used to test the effect of operator deviations from assigned departure times on performance.

Chapter 7 summarizes the findings of this research, and presents recommendations for operational and policy changes at the MBTA. In addition, potential avenues for future research are described.

Chapter 2

Literature Review

This chapter reviews the literature on transit service reliability and bus operations control, including both models and experiments. The literature informs the direction of this thesis in examining holding strategies and skip-stop strategies using both experimental and model data. Section 2.1 covers literature on transit service reliability, including measures of reliability, causes of unreliability, supervision practices, and data collection methods.

2.1 Transit service reliability

Abkowitz et al. (1978) presented a comprehensive review of transit service reliability, including the impact of reliability on passengers and on the transit agency, empirical measures of reliability, causes of reliability problems, and techniques for improving reliability.

The authors noted that reducing travel time and wait time variability increases the utility of transit relative to other modes, and can attract new riders to transit as well as increasing transit use by existing riders. They also noted that reliability improvements can reduce capital and operating costs for agencies.

Causes of unreliability were grouped into "environmental" and "inherent", with traffic and demand variability being two of the most significant causes of unreliability.

Methods for improving reliability were classified as priority (signal priority or lane priority), control (holding or skip-stop strategies) and operational strategies (improvements in timetables or fleet and labor management).

By enumerating and classifying the most common causes of unreliability, Abkowitz et al provide a foundation for most subsequent research in the area of unreliability.

2.1.1 Bus supervision

Levinson (1991) conducted a review of bus supervision practices used by twenty North American transit agencies. The review identified four main factors that contribute to reliable bus service:

- Realistic routes and schedules
- Adequate maintenance
- Sound personnel policy
- Effective supervision

Levinson identified several impediments to good supervision on bus routes, mainly tracing back to limited financial resources. Financial difficulties led to insufficient numbers of supervisors, as well as increased maintenance problems which took away supervisors' time from monitoring service. Inadequate communications technology was another impediment to proper supervision.

Pangilinan (2006) created a framework for the deployment of supervisory personnel to manage and improve reliability of bus service. The study examined the roles of three types of supervisory personnel: Post supervisors, mobile supervisors, and control-center staff. Based on an assessment of the availability of information, communications, and personnel, recommendations are made for system-level deployment of personnel. Case studies are performed on Chicago Transit Authority Route 20 and the MBTA Silver Line Washington Street route.

Pangilinan's overall recommendations involved changing supervisor roles as better communications technology becomes available; for example, moving staff from post-supervisor to mobile-supervisor roles. Eventually, Pangilinan recommended a focus on terminal departure adherence with automated instructions given to operators similar to a "ring-off bell", as well as "exception-based reporting" in which AVL data is processed in real time to draw the attention of control-center staff to large gaps or other bus reliability problems.

2.1.2 Measuring unreliability and its causes

Cham (2006) developed a framework for applying automated data collection to the evaluation of service reliability, first using metrics to measure reliability, and then determining the causes of unreliability. The framework was used to examine the MBTA Silver Line Washington Street bus service, and Cham determined that irregular departure times from the terminal were the main cause of unreliability. Trips that departed the terminal with leading headways close to the scheduled headway were much more likely to maintain that headway downstream than trips that left with larger or smaller headways. Cham's recommendations for the Silver Line include improved terminal supervision as well as signal priority and increased separation of the right-of-way. Her paper motivates our focus on terminal departures in this thesis.

2.2 Automatically-collected data for transit service improvement

Research on uses of automatically-collected data for transit service improvement is summarized by Furth et al (2006). The authors review the uses of archived AVL and APC data for service planning, scheduling, and performance evaluation. They describe the historical uses of AVL data for real-time applications, and APC data for after-the-fact analysis. In particular, they note that the typical AVL system does not provide archived data in a useful format for analysis.

Furth et al also discuss the various providers of software that may be used for analysis of archived data: transit agencies (in-house software), equipment vendors, scheduling system vendors, third-party vendors, and researchers. They note that in-house software has the most flexibility but a high cost in staffing and maintenance, while software from third-party vendors has flexibility but may be difficult to justify within the constraints of a typical transit agency budget.

The authors generally conclude that the transition to a data-rich environment provides opportunities to expand the analysis of performance, including setting new, more precise service standards. The need for integration with related databases, such as stop locations, schedule information, and fare-collection data, is also highlighted. This study, although it does not consider real-time decision tools, provides a useful analysis of the choices available to agencies during the software procurement process, as well as the types of data available to agencies. These concepts will be utilized in this thesis in the context of an automated decision tool.

2.3 Simulation models of bus service

Moses (2005) created a simulation model of a CTA bus route, but was unable to validate the model. The main difference between the simulation and real AVL data was that the simulation had more irregularity than was observed in reality. Varying parameters such as standard deviations of travel times and passenger demand levels did not solve the problem. Moses suggested two reasons for the lack of validation: correlations between parameters such as successive run times or passenger demand levels, and operator behavior such as purposely slowing down or speeding up to even out headways.

Milkovits (2008) developed and validated a simulation model of CTA Route 63, adding to previous models a detailed treatment of schedule deviation at terminal departures. By explicitly modeling terminal departure behavior, Milkovits was able to accurately recreate the conditions on the route.

Sanchez-Martinez (2012) developed a simulation model of a high-frequency bus route in London. It was used for the purpose of testing allocation of resources on the route. The key addition of this model was the use of a bivariate running-time distribution, in which running times on route segments were randomly drawn from a distribution of observed vehicle running times, grouped by two factors: the time of day, and the vehicle's running time on the previous segment.

2.4 Heuristic strategies for transit control

Several authors develop insight into control strategies through exploration of heuristic rules that either provide approximate solutions to intractable optimization problems, or exact solutions to simplified versions of these problems. These rules take as inputs factors such as vehicle locations, running times, and arrival rates, and output values that define control actions, such as holding time at stops.

2.4.1 Holding

Turnquist (1981) examined vehicle holding strategies, including schedule-based holding for low-frequency routes and headway-based holding for high-frequency routes. He compared two possible holding strategies: The "Prefol" strategy, in which a vehicle is held to split the headway between the preceding vehicle and the following vehicle, and the "scheduled headway" strategy, in which each vehicle is held until the scheduled headway has elapsed since the previous vehicle's departure. Using a simulation model, Turnquist found that the Prefol strategy is superior to the scheduled-headway strategy, although it loses its advantage as successive headways become more strongly correlated.

Turnquist and Blume (1980) used a probabilistic model to identify situations where holding controls are effective. In particular, they noted that the ideal control point is one where relatively few people are on the vehicle and many passengers are waiting to board at subsequent stops. This maximizes the benefits (which accrue to downstream passengers who experience more regular headways) while minimizing the costs (which mainly fall on passengers already on a bus that is held). In general, these criteria lead to the ideal control point being at or near the departure terminal.

2.4.2 Skip-stop strategies

Skip-stop strategies are strategies in which vehicles skip some of their scheduled stops. These include deadheading, expressing, and short-turning. The costs and benefits of these strategies are similar, as each involves a trade-off between additional wait time for skipped passengers, and saved time for downstream passengers.

Eberlein et al (1998, 1999, 2001) examined holding, deadheading, and expressing strategies using a deterministic quadratic program with a rolling horizon. The model has the objective of minimizing total passenger wait time, and includes the effect of dwell time on vehicle delay and headways. A heuristic solution to the combined control problem of using all three strategies is presented, and tested in a simulation. Eberlein found that the use of both holding and stop-skipping strategies resulted in improved performance over a single strategy, as well as reducing the frequency and extent of stop-skipping.

Song (1998) developed a heuristic strategy for controlling rail service (using the MBTA Red Line as a case study) from a terminal, including holding, expressing, deadheading, and short-turning possibilities. The author uses a model with an objective of minimizing passenger waiting time, with deterministic travel times and dwell times modeled as a function of the number of passengers boarding and alighting from the train.

2.5 Rolling-horizon optimization

In addition to heuristic strategies, rolling-horizon optimizations are another method for making control decisions. These involve predicting how service on a route will continue over a limited time horizon, and using optimization techniques to search through a constrained set of solutions. With continuing advances in computing power, real-time use of these optimization techniques has become feasible.

Delgado et al (2012) created a non-linear model for optimization of holding times of all vehicles of a transit line at all stops, taking into account vehicle capacity constraints. In addition to holding, limiting passenger boardings was also allowed by the model. The model assumed passenger arrival rates and running times between stops were constant over time, and it was tested in a simulation.

Sanchez-Martinez (2014) built upon the model of Delgado et al, adding running times and passenger arrival rates that vary dynamically over time. Sanchez-Martinez found that holding controls based on optimization with these dynamic inputs were superior to those based on static inputs, as well as to the "even headway" or "threshold headway" heuristic strategies. The improvement was found to be largest under conditions of heavy crowding, and limited under low to moderate crowding conditions.

2.6 Holding experiments

Experiments with bus holding strategies began decades ago, but prior to the availability of vehicle location data in real time, they required large numbers of staff, both to execute the strategies and to collect data. With the advent of AVL systems that provided vehicle locations in real time, as well as advances in mobile device technology, a broader variety of experiments became feasible. In this section, experiments are described that span a variety of different strategies and implementation methods.

Abkowitz and Lepofsky (1990) performed an experiment on two MBTA bus routes: Route 1, a crosstown route, and Route 57, a radial route. A threshold-based holding strategy was applied, in which buses were held until a minimum headway was reached from the previous bus. Various points including both midpoints and terminals were used for the control strategy. They found a small but noticeable improvement in headway and travel-time reliability when Route 1 was controlled at a midpoint, and no improvement when Route

1 was controlled at Harvard or when Route 57 was controlled at a midpoint. Their results were limited by small sample sizes and difficulties with manual data collection and manual implementation of the strategy.

Strathman et al (2001) conducted an experiment using real-time AVL data to inform supervision of buses departing downtown Portland during the PM Peak period. A dispatcher using the AVL system communicated with a supervisor who would instruct buses to depart based on Turnquist's "Prefol" strategy of splitting headways. Other strategies available to the supervisor included short-turning and substitution of runs in the schedule. The results showed that headway variances declined 3.8% overall and 15.8% at the control point, with most of the benefits appearing at the first three timepoints on each route. Passenger load variance also decreased by 16%, and the authors conclude that "small improvements in service regularity can potentially generate more substantial improvements in passenger load maintenance."

On CTA Route 20, Pangilinan et al (2008) tested a prefol strategy implemented at a terminal and two points along the route, in the AM Peak period in the peak direction. Allowed control actions were: holding, "dragging the street" or driving the bus more slowly than usual, and departing earlier than scheduled from the terminal. The supervisors on the street were in communication with a control-center supervisor who gave them instructions based on real-time AVL data. The realized reduction in variation from the experiment was less than that predicted by a Monte Carlo simulation, but its effects persisted farther down the route than in the simulation.

Problems with implementation of the controls included:

1. The control-center dispatcher not being able to devote full attention to the experiment, and thus missing some big gaps

2. Missing real-time data, when some buses did not appear in the AVL system

3. Variations in departure time caused by factors at the terminal, such as passenger boarding, distractions, and other tasks performed by the on-street supervisor.

At the mid-route control points, holding more than 1-2 minutes was deemed infeasible due to irritation caused to passengers already on-board the bus.

Bartholdi and Eisenstein (2012) created a new method for holding buses, based solely on the trailing headway of the control vehicle, that is, the headway between the control vehicle and the next vehicle approaching the control point. The amount of time to hold a vehicle is given as αh_n, where h_n is the trailing headway and $0 < \alpha < 1$ is a control parameter that determines sensitivity to perturbations. The authors first verify using an idealized model that this will lead to more regular headways, then conduct an experiment on a Georgia Tech shuttle bus route through Atlanta using control points at the ends of the route. They find that the strategy leads to reduced variation in headways and elimination of severe bunching. They also find that it responds quickly to the removal of a bus from service.

Xuan et al (2011) formulated a holding strategy based on a "virtual schedule" of predicted arrival times at stops (as opposed to the published schedule which typically shifts times earlier to avoid early departures from stops). They use deviations from this virtual schedule of the control vehicle and the trailing vehicle as inputs to determine holding times at points along the route. They find that this strategy improves both headway regularity and schedule adherence.

The authors have since founded a start-up company called VIA Analytics, which has implemented a version of this control strategy using tablet computers placed on-board buses, giving instructions directly to bus drivers. They have installed their system on two bus routes in San Sebastian, Spain, and found that excess passenger wait times decreased by 40%, along with reductions in schedule deviation (VIA Analytics, 2013).

Finally, Lizana et al (2014) tested the optimization-based strategy developed by Delgado et al (2012), which was described in Section 2.5. They used tablet computers in a similar fashion to VIA Analytics, and observed reductions in wait time and crowding when the strategy was implemented on two bus routes in Santiago, Chile.

2.7 Summary of literature review

Abkowitz et al, Levinson, and Pangilinan summarized measures taken to improve bus service reliability, including priority (signal priority or exclusive lanes), control strategies, and operational strategies (timetables and fleet management). Good communication, lack of distractions from other duties, and appropriate positioning of supervisors were found to be key to effective controlling. This thesis will further explore the factors that cause difficulty with precise controls of bus departure times.

Many different models have been used to test different holding strategies. Heuristics such as Turnquist's Prefol strategy and Bartholdi and Eisenstein's method based on trailing headways have held up well under simulation and experimental testing. Rolling-horizon optimization routines such as those created by Delgado and Sanchez-Martinez are able to improve upon the heuristics, mainly in cases of heavy crowding.

Experiments with bus control strategies have been performed by Strathman et al, Bartholdi and Eisenstein, and Xuan et al, among others. A variety of heuristic and optimization strategies have found success, although communications problems, operator compliance, and poor data have appeared as difficulties. If vehicles are held at midpoints, the experiments typically limit this holding to a maximum of 1-2 minutes to avoid passenger irritation. The experimental portion of this thesis will use a terminal-based control strategy, similar to the work of Strathman et al and Bartholdi and Eisenstein.

Simulation models have been used extensively in the study of control strategies, such as by Delgado et al and Sanchez-Martinez. This thesis will use a simulation model based upon the work of Sanchez-Martinez to test the effects of changing terminal departure behavior, as well as midpoint-holding and short-turning strategies.

Chapter 3

Automated tools for transit performance improvement

Automatically-collected data from transit vehicles has been used to manage and improve service for many years, both through real-time tracking of vehicles, as well as archived data for later analysis. Recently, technological advancements have enabled a new application of real-time data: Automated decision-support tools. These can both improve performance directly through real-time control of service, and also improve data analysis by adding explicit information about what control actions were taken.

In this chapter, we will show how automated decision-support tools can be created to improve both management of service and analysis of performance. We will first define what a decision-support tool is, and describe how it fits into the process of improving performance and operational control. Next we will discuss the various control strategies that can be implemented using automated tools, and the requirements for their implementation. We then describe the way that archived data from an automated decision tool can improve the analysis of performance and support service planning, by giving analysts an accurate, automatically-collected record of when control actions are taken. Finally, we present a generic design for a decision support tool based on automatically-collected data, followed by a specific case study consisting of a tool developed to implement a holding strategy on an MBTA bus route.

3.1 Automated decision-support tools

In the context of bus transit operations, an automated decision-support tool is software that uses real-time vehicle location data to provide recommendations for control actions. These may come in a variety of forms, from exception-based suggestions made to a dispatcher that can be followed or ignored, to departure-time instructions given directly to a bus operator trained to follow the automated directions.

Furth et al (2006) describe the use of archived Automatic Vehicle Location (AVL) and Automatic Passenger Counter (APC) data to improve transit performance and management. They place these automatically-collected data in the context of "service quality improvement cycles" (Figure 3-1), showing how these data sources can improve both real-time operations management and passenger information (the "real time loop") and the analysis of performance and demand (the "off-line loop").

In Figure 3-2, we show how an automated decision-support tool can fit into Furth's service-quality improvement framework. It contributes directly to operational control by providing control decisions, but it is also a key contributor to the off-line data used for performance and demand analysis. It enables analysts to directly observe whether, for example, a holding strategy was used, when analyzing factors such as on-time performance or dwell time. Without the automated tool, either a manual record of control actions would have to be kept, or analysts would need to infer control actions based on AVL or other data. Real-time control will be discussed in Section 3.2 and uses of off-line data in Section 3.3.

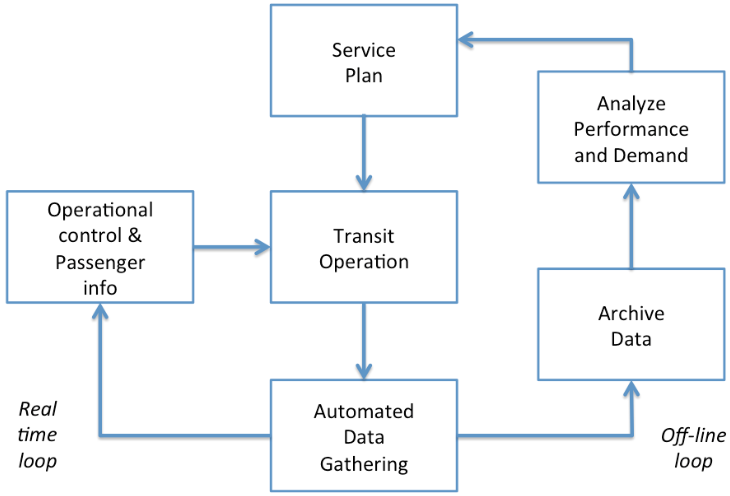

Figure 3-1: Service quality improvement cycles (Furth et al, 2006)

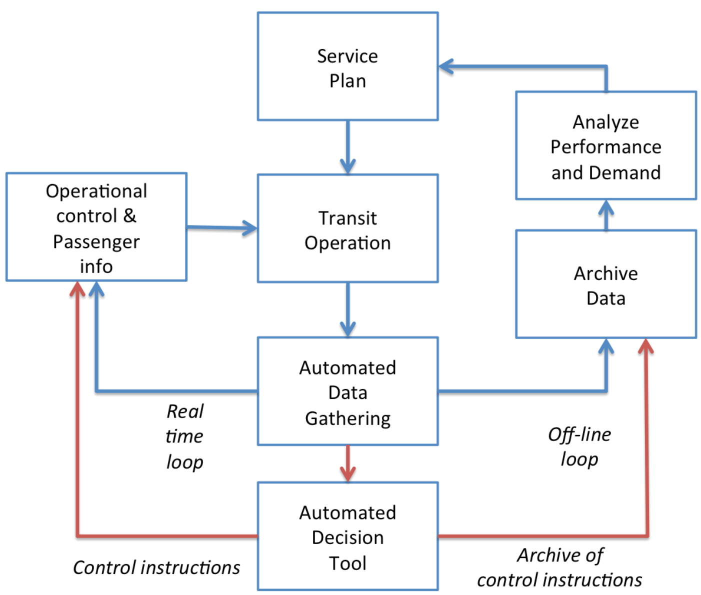

Figure 3-2: Service quality improvement with automated decision tool

3.2 Real-time control strategies

Implementing more effective real-time control strategies is the immediate motivation for the creation of decision-support tools for transit. Such strategies can be applied without any automatically collected data, but the use of automation reduces the amount of resources required to implement them, and increases the effectiveness of the resulting decisions. In this section, we will discuss holding, deadheading, expressing, and short-turning strategies, and how an automated decision-tool could be used to implement them or to improve existing implementations.

3.2.1 Holding

Holding is the strategy of instructing an operator to remain at a stop for a period of time to improve schedule adherence or the spacing between vehicles. A variety of holding strategies exist, including schedule-based, headway-based and optimization-based strategies. The strategies can be applied at any number of control points, from a single timepoint up to every stop on the route.

Schedule-based holding entails delaying buses that arrive early at control points until their scheduled departure time. Implementing this strategy requires schedules with a well-adjusted amount of slack time, along with supervision to ensure that buses depart on time (Turnquist 1981). Too much slack time leads to buses spending too much time idling, causing increased travel time for passengers. Too little slack time results in buses running late most of the time and the strategy not being applied.

Schedule-based holding can be implemented without the use of any AVL data. It requires only that bus operators be aware of the scheduled departure time from each control point. Because of this, schedule-based holding is a popular control strategy among transit agencies. However, the strategy relies on operator compliance, particularly when holding is to be applied mid-route without a supervisor present. In this respect, AVL data is very helpful because it allows monitoring of many operators by a single supervisor, either in real-time or after the fact with archived data.

Headway-based holding refers to any strategy in which holding decisions are made based on the headways of buses on the route. This recognizes the fact that on high-frequency routes, passengers generally arrive at stops without regard for the schedule. Assuming that passengers arrive randomly, the expected wait time per passenger is a function of the distribution of headways, and increases with the square of the coefficient of variation of headways. This motivates the focus on achieving even headways over schedule adherence on high-frequency routes.

Approaches to headway-based holding range in what factors they take into account, and thus what types of data they require. Various approaches that have been tested in the literature are described below.

- The "Single Headway" strategy (Turnquist, 1982): Hold a bus until its preceding headway is equal to a defined minimum headway. Required data: Departure time of the preceding bus.

- The "Prefol" strategy (Turnquist, 1982): Hold a bus until its preceding and trailing headways are similar. Required data: Departure time of the preceding bus, predicted departure time of the trailing bus.

- The "Self-equalizing headway" strategy (Bartholdi and Eisenstein, 2012): Hold a bus for a length of time equal to its trailing headway multiplied by a control parameter $0 < \alpha < 1$. Required data: Predicted departure time of the trailing bus.

The headway-based holding strategies all rely on the actual departure times of the preceding bus and/or the predicted departure time of the trailing bus. The departure time of the preceding bus can be observed without AVL data, by a supervisor posted at the control point, but this is a very inefficient use of personnel, and vulnerable to measurement error or inattentiveness by the supervisor. In any realistic application, AVL data must be used to calculate holding times. Bartholdi and Eisenstein note that the predicted departure time of the trailing bus is easier to obtain from publicly-available data sources than the departure time of the preceding bus. This will be discussed further in Section 3.4.2.

Rolling-horizon optimization is a set of strategies that have been developed much more recently, as they require real-time vehicle location data and significant computing power. The basic framework for such a strategy requires an optimization model, a performance model, and a cost model. The optimization model feeds information about the system (running time, demand, locations of vehicles) as well as potential holding times into the performance model. The performance model predicts the evolution of the system over a "rolling horizon" and feeds this information into the cost model, which calculates the expected cost (Delgado, 2012, and Sanchez-Martinez, 2014). The costs are then used by the optimization model to select the optimal holding times.

This general framework allows for a variety of real-time or static data to be used as inputs, depending on the implementation. Information about the locations of vehicles is required at a minimum, while data on passenger demand and running times could be calculated in real time or based on historical observations.

3.2.2 Deadheading and expressing

Expressing and deadheading are two strategies that involve skipping stops on a route, to improve spacing. Expressing involves sending a bus that currently has passengers on-board to a downstream stop bypassing intermediate stops. It involves a trade-off between two sets of passengers:

- Negatively impacted passengers: Those on-board whose destinations are skipped (who must alight and transfer), and those waiting at skipped stops downstream.
- Positively impacted passengers: Those on-board whose destinations are beyond the express segment, and those waiting beyond the express segment (who may see reduced waiting times).

The ideal situation for expressing a vehicle is one in which the preceding headway is long and the trailing headway short, and passenger demand beyond the express segment is high (Wilson et al, 1992). This scenario maximizes the cost-benefit ratio to the impacted passengers.

Deadheading is similar to expressing, but involves taking a bus out of service, typically at a terminal, and running it empty over a segment of the route. The groups of passengers impacted are then simply those waiting downstream who are skipped, and those beyond the deadhead segment, who may see reduced waiting times. This is generally only done at terminals because deadheading from a midpoint would force all passengers to alight early, and thus be strictly worse than expressing from a midpoint (as some passengers could remain on-board). Deadheading saves time over expressing because passengers are not boarding at the initial stop.

The availability of real-time data greatly enhances the ability of agencies to apply deadheading and expressing strategies. The data required to evaluate a potential deadheading or expressing action are leading and trailing headways, as well as passenger demand. Vehicle locations and predicted times at stops are readily available via AVL systems, as discussed in the previous section. Passenger demand at downstream stops, on the other hand, must be estimated. This can be done using historical demand patterns combined with knowledge of recent headways. If APC or AFC data are available in real time, these can be valuable inputs to the estimation of demand along the route.

3.2.3 Short-turning

Short-turning involves ending a trip early, i.e. at a stop prior to the terminal, and immediately beginning the next trip in the reverse direction. Short-turning is ideally applied to a bus when the passenger load is small, the following headway is small, and there is a large gap in service in the reverse direction (Wilson et al, 1992). The groups of passengers affected by a short-turning strategy are very similar to those affected by expressing: A segment of the route is not served, to the benefit of service downstream of the segment. The main difference from the other stop-skipping strategies is that with short-turning, passengers in the skipped segment do not see a bus pass them by, and likely are unaware that any control action has been taken.

The real-time data requirements for a short-turning decision tool are similar to those for expressing and deadheading. In this case, rather than downstream demand, the passenger load on the vehicle is the key variable not available via AVL data. This can be estimated based on historical demand patterns and preceding headways, or, if it is available in real time, APC or AFC data can be used.

3.3 Performance analysis and service planning

In addition to enabling implementation of control strategies to improve service, an automated decision-support tool can provide benefits in the form of the data it archives. In this section, we discuss the benefits to performance analysis and service planning that are derived from the archived data.

3.3.1 On-time performance

The analysis of on-time performance is crucial to service planning. A late departure from a stop may be caused by insufficient scheduled time, poor operator behavior, control actions applied by a supervisor, or external factors such as heavy demand at the stop. An early departure could also be caused by either poor operator behavior or a deliberate control action. In order to improve on-time performance, the causes of these irregular departures must be identified.

Cham (2006) proposed a method to infer the causes of poor on-time performance at terminals based on the recovery time available to an operator on arrival at the terminal and several assumptions about the minimum required dwell time at the terminal and when supervisors might take deliberate actions affecting departure time. It is a useful approach, but its heavy reliance on assumptions leads to uncertainty and would likely be unpopular as a method of identifying poor on-time performance by individual operators.

The use of automated decision tools removes the need to infer whether control actions were applied. By examining archived data from an automated decision tool, managers can determine much more accurately whether early or late departures were caused by control actions. This reduces the uncertainty of who is responsible for individual early departures, paving the way for operator-specific discipline or interventions. Late departures may still be caused by factors beyond the control of the operator, and must still be examined carefully.

3.3.2 Running-time analysis

Running-time analysis is a key part of any scheduling process. Furth et al (2006) describe the data agencies use to set schedules. Agencies frequently set scheduled running times and half-cycle times for routes based on percentiles of observed running time. Agencies that use a schedule-based holding strategy, very common in North America, face a challenge in identifying the running-time distribution to use to set their schedule: They must exclude time spent holding at stops, or else their estimates will be biased.

Some agencies examine data on when doors were open; others use unusually long dwell times as a signal that holding may have occurred. When using an automated decision tool, identifying time spent holding should be easier, regardless of what holding strategy was used.

3.4 Decision-support tool for real-time control of high-frequency bus routes

In this section, we describe a design for a mobile phone or tablet application that uses real-time bus location data to provide decision support for operations control. In the following section, we will describe a specific implementation of this design. It is specifically oriented toward decision rules that are fully automated; that is, they take as inputs data about the current (and recent) state of the system, and output a recommended control action. The only decision made by a human is whether or not to follow the application's recommendation. The components of the software application are shown in Figure 3-3.

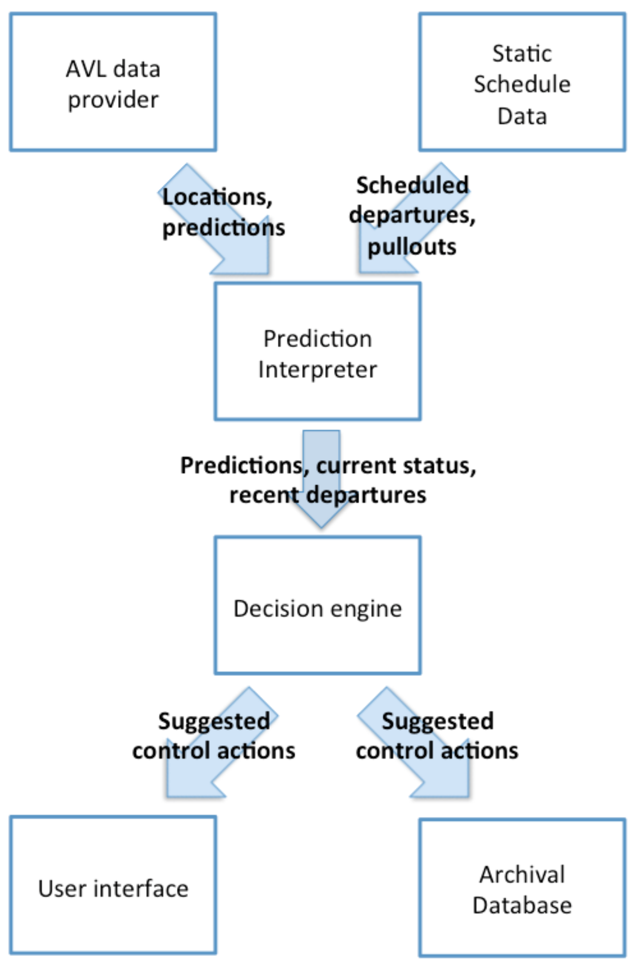

Figure 3-3: Framework for decision-support software

3.4.1 Data sources

The data required for real-time control of buses is, at a minimum, locations of all buses on the route. Other data provided by a typical CAD/AVL system include identifiers for the vehicle, block, and trip, as well as scheduled and predicted arrival/departure times for downstream stops. In a software application context, the physical source of the data (AVL hardware) is not the specific concern, as we assume that the agency involved has already procured this equipment. Rather, the important factor is the Application Programming Interface (API) through which the data is downloaded. APIs available for transit vehicle location and prediction data differ across several dimensions:

- **Public availability** - Using data from a feed that is open to the public provides a more convenient base from which to program applications (due to fewer security requirements) and reduces the maintenance needs of the agency (which does not need to maintain two separate feeds). A feed for internal use only, on the other hand, may include more data, such as operator IDs and operator run information, which agencies may not wish to include in the public feed.

- **Standardization of format** - Some transit data feeds use standardized formats such as GTFS-realtime, which allow apps written for one agency to be easily ported to other agencies' data feeds. Others find that the existing standardized formats do not meet all their needs, and create proprietary formats, which may be based on an existing standard but with additional features added, or may be entirely proprietary.

- **Ownership of data feed** - The implementation of AVL data feeds is typically done through a contract with an information technology company rather than by agency staff (with New York City Transit being a notable exception). Depending on the details of the contract, the agency may have full control over the data and the data feed, or the contractor may retain various rights, such as the copyright over a proprietary format, or even the right to exclusive use of the data that is provided.

- **Data downloading process** - Some data feeds, such as GTFS-realtime, require all data from an agency to be downloaded at once. This is ideal for an application with heavy data needs, across many different routes. Such an application must host the entire dataset on a server and periodically download it in full. Other data feeds use “web service” interfaces that allow only the data specifically needed by a single user to be requested. These are appropriate for more limited use cases.

Each of these factors is important to consider in any implementation. In many cases, agencies will have multiple available data feeds for AVL data, and application developers can select the most appropriate for a particular use case.

3.4.2 Prediction interpreter

Typically, AVL data feeds provide two basic types of data: Locations and predictions. Locations typically consist of latitude-longitude coordinates from the most recent observation of the vehicle, while predictions typically give a projected arrival time for at least one downstream stop. As described in Section 3.2.1, many control strategies require as inputs not only predicted arrival times, but also a recent history of departure times from stops previously visited by each vehicle.

For this application, we develop a software component which we refer to as the “prediction interpreter”, which takes location and prediction information from the data feed, and translates it into a useful data set for the application of the control strategy, including recent departure times. The logic for this component will depend heavily on the data feed used, but the basic concept consists of observing when a prediction for a vehicle’s arrival at or departure from a stop disappears from the feed. This indicates that the vehicle has arrived at (or departed from) the stop. If this method is found to be unreliable (which commonly happens at terminals), then an additional criterion can be added, specifying a minimum distance from the stop that a vehicle must report in order to be considered to have departed. The specific implementation must be developed using knowledge of the data feed and testing against human observations.

3.4.3 Decision engine

The other key component of the system is the decision engine, which creates suggested departure times for vehicles based on a given rule. It must accomplish the following:

1. Read in current status of vehicles, including predicted arrival times, recent departure times, and locations, from the prediction interpreter module

2. Apply the decision rule to determine the suggested departure times for vehicles to be controlled.

3. Send all data about the current state of the system to the database module.

4. Send all data about the current state of the system to the user-interface module.

The decision engine implements the control-strategy logic. In many cases (when no additional information is needed), it may be the only module that needs to be changed to implement a new strategy. As seen in Section 3.2, many control strategies can be implemented with a full set of vehicle locations and predictions on the route. In a situation where APC or AFC data were available in real time, the decision engine could be modified to read in those data as well as inputs to the control strategy.

3.4.4 User interface

The user interface must be adapted for the type of user and the type of device used to view the app. There are three basic categories of operations personnel who might be users of a real-time decision-tool:

- Dispatchers using a desktop-computer interface. They often face a heavy workload and would likely be best-served by an "exception-based" system such as that described by Pangilinan et al (2006). This would alert the user only if a threshold of importance is reached, which could be calibrated based on the dispatcher's workload.

- Supervisors may use a tablet or mobile-phone to access the app, or have a laptop if they are assigned a vehicle. Location-based services could be helpful in this case, as supervisors are likely close to the vehicles they are supervising. Bartholdi and Eisenstein (2012) used this type of app.

- Operators must access an app through a tablet or touch-screen built in to their vehicles. An app aimed at operators must have an extremely simplified user interface, to avoid distracting the operator. Lizana et al (2014) as well as the start-up company VIA Analytics have both deployed this type of app.

3.4.5 Archived data

In addition to automating the application of a control strategy, a decision tool also provides a new source of automatically-collected data for performance evaluation and service planning. Having a record of the control actions that were recommended by the app adds value to the existing AVL dataset and allows for more accurate and more detailed analysis of bus movements. These uses were described in Section 3.3, and here we describe the implementation of the data archive.

The archive should be stored in a relational database, ideally as part of a larger system already present at the transit agency. Since the use of a tool like this requires that an agency already have at least an AVL system, it can be assumed that some type of database exists. It is crucial to integrate the new data source on control decisions into the existing database structure to allow for the new information to be easily integrated into existing analytical systems. As Furth et al (2006) note, the utility of transit datasets depends on their effective integration.

3.5 Implementation for MBTA experiment

This section describes a particular implementation of the decision-support tool design from Section 3.4. The implementation is a mobile app used to provide instructions to MBTA bus supervisors based on Turnquist's "prefol" strategy described in Chapter 2. Chapter 4 describes an experiment performed using this app on MBTA Route 1. Here we describe how each component of the design is implemented.

3.5.1 Data sources

For this application, we chose to use publicly-available data feeds. These are easy to access, require no special permission, and can easily serve the needs of a decision-support tool with only minor modifications. The MBTA provides AVL data to the public in three formats: GTFS-realtime, the NextBus API, and MBTA-realtime. Their characteristics are described in Table 3.1.

GTFS-realtime is an extension of the General Transit Feed Specification (GTFS), an open standard for public transportation data popular among North American agencies. It has the advantage of being a popular open standard, so that code developed to process a

Table 3.1: Real-time data feeds

Data feed	Modes	Requests available	Agencies using	Owner of feed
GTFS-realtime	All modes	Full system dataset	Many agencies	MBTA
NextBus API	Bus only	Specific queries	Many agencies	NextBus
MBTA-realtime	All modes	Specific queries	MBTA only	MBTA

GTFS-realtime feed from one agency should be easily usable with data from another agency. It is designed in such a way that the data for an entire agency must be downloaded at once, which is a benefit in systems aiming for efficiency in a large-scale implementation, but a negative for applications which require only a small subset of the data.

NextBus is a company that provides arrival predictions for transit vehicles based on AVL data. The company is contracted by the MBTA to provide predictions to the public via its web application as well as its open API. The advantage of the NextBus feed is the accuracy of its predictions, which are based on more sophisticated algorithms than the simple lateness measure used by the MBTA's CAD/AVL system, and the simplicity of its API. The main disadvantage is that it is a proprietary data feed, and if the MBTA ceased its relationship with NextBus, the company would probably not continue providing the feed, as it does not make any advertising revenue from its website.

Finally, MBTA-realtime combines features of the other two. It is a web API, like NextBus, and thus provides flexibility for developers to request only the subset of the data that they need. Unlike NextBus, its source code is owned by the MBTA, and therefore there is little reason to worry that it might disappear. Similarly to GTFS-realtime, it suffers from the fact that its predictions are simply based on the estimated lateness at the current location from the CAD/AVL system.

For our app, we decided to use the NextBus API because of its more accurate predictions and ease of use. However, it shares a problem with all three data feeds: As a passenger-oriented service, it focuses on predicting future arrival times of buses, and does not provide any history of departure times, which is necessary to execute the prefol strategy. Methods of dealing with this issue are discussed in the next section.

3.5.2 Prediction interpreter

The prediction interpreter is the critical component to transforming the passenger-oriented predictions feed into a useful dataset for control strategies. For the prefol strategy, we need the most recent departure time of a vehicle from the control point. Our prediction interpreter accomplishes this by downloading a set of predictions for the terminal stop and several nearby stops, as well as the location of the next bus to depart. Because the API provides only predictions, and these predictions disappear as soon as a bus has departed, we infer whether or not a bus has departed based on whether it has a prediction available. Because NextBus was observed occasionally to remove buses from the departure predictions even when they had not departed, we also set a minimum distance from the terminal of 75 meters. This threshold value was determined through observations; over two hours of observations at each terminal, 75 meters was found to be sufficient in all observed cases to distinguish between buses which had left the terminal, and those which remained at the terminal but whose predicted departure time had disappeared for some other reason. The steps taken by the prediction interpreter are described below. They are executed every 15 seconds, after downloading the most recent data from NextBus (although each individual bus transmits a location update every 60 seconds).

1. Download predictions for the terminal stop and all downstream stops
2. Based on whether or not arrival and departure predictions exist for each bus at the terminal, divide buses into three categories: "Approaching" (prediction exists for arrival at terminal), "At Terminal" (prediction exists for *departure* from terminal), and "Departed" (only predictions for downstream stops exist).
3. If a bus previously categorized as "At Terminal" has moved to the "Departed" category, check if its location is at least 75 meters away from the terminal.
   (a) If the bus is at least 75 meters away from the terminal, estimate its departure time from the terminal as follows: Given the predicted arrival time t_s at downstream stop s (the first stop for which predictions are available), the current time t_0, and the average running time r_s from the terminal to stop s, estimate the departure time as $\min(t_s - r_s, t_0)$.
   (b) If the bus is less than 75 meters away from the terminal, re-categorize it as "At Terminal."

In the worst-case scenario, a bus will be mis-categorized as "At Terminal" when it has in fact departed but not yet traveled 75 meters, but this situation is unlikely to persist for any significant period of time, as it will be corrected after bus reaches this threshold distance. The method of estimating departure times is crude, as it uses the average running time over all peak and midday time periods, but because the stops used are very close together (typically the first stop after the terminal is the one for which predictions are used in step 3a) the running times on the segment do not vary greatly. This method was tested against in-person observations over three hours during peak and midday periods at Harvard and Dudley; the root mean squared error (RMSE) was 9 seconds at Harvard and 12 seconds at Dudley, and the 95th percentile deviation was 19 seconds at Harvard and 18 seconds at Dudley.

3.5.3 Decision engine

In our case the decision engine includes the prefol strategy with a constraint that no vehicle may depart earlier than scheduled. The decision engine consists of the code defining the strategy's logic, and the code that interfaces with the input and output modules. Although it is central to the overall structure of the application, it requires less code and is generally less complicated than the other modules, due to the simplicity of the strategy being implemented.

3.5.4 User interface

The initial decision made for the user interface was who the target user would be. Due to resource limitations, it was decided that the holding strategy would be implemented by supervisors stationed at terminals, using either a mobile phone or handheld computer. After several iterations, including a test run where a supervisor used the app to control bus departures from one terminal, we settled on the user interface shown in Figure 3-4.

The main difficulty in designing the UI was in striking a balance between providing more information about buses on the route, and simplifying to show at a glance whether, and for how long, to hold the current bus at the terminal. We decided to show the predicted arrival times of the next two upcoming buses, as well as the scheduled, suggested, and actual departure times of buses that departed recently, to provide context for the supervisor to understand why the app was making its recommendations, and to provide a reminder that data on adherence to the app suggestions was being recorded.

Figure 3-5 shows the three possible types of messages that could be given by the app. The larger, yellow typeface indicates that the user must take an action, while the smaller, green typeface indicates that no special action is required.

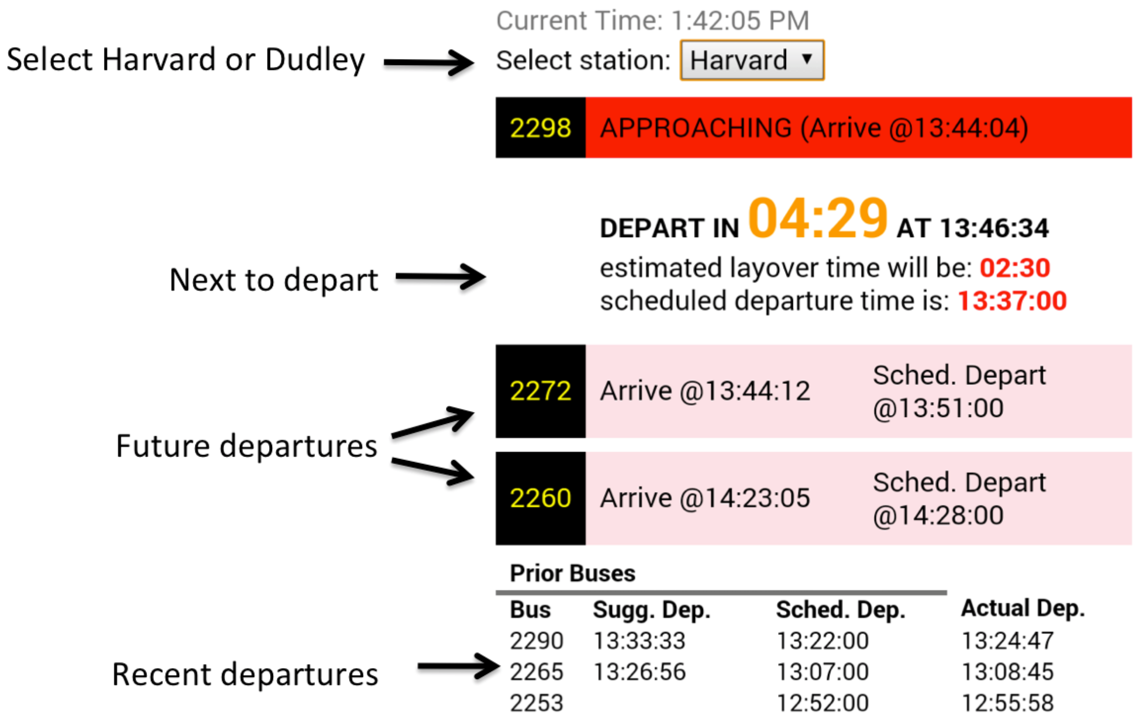

Figure 3-4: Screen shot of app, taken on an Android phone running Google Chrome

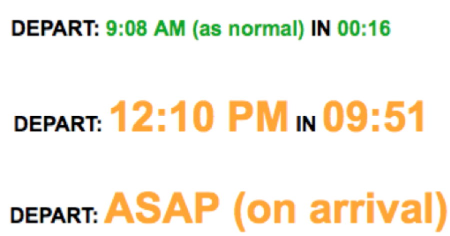

Figure 3-5: Types of instructions, from top: depart on schedule, depart after holding, depart as soon as possible

3.5.5 Archived data

In our app, the decision engine writes records to a single database table, called "snapshots." The snapshots table records the state of the system every time the decision engine refreshes, including details on the status of all vehicles as provided by the AVL system, along with the suggested departure time provided by the decision engine. Vehicles in this table are identified by vehicle ID, operator ID, and trip ID, which is sufficient to match the snapshots to records in the AVL timepoint database.

The table is stored in the "data warehouse", a PostgreSQL database created by MIT researchers to unify the various automatically-collected datasets available at the MBTA. Storing the data in this database makes it easy to join data from, for example, the APC or AFC systems to the decision-tool records.

3.6 Summary

In this chapter, we have outlined the components of an automated decision-support tool, the decisions that must be made in implementing such a tool, and its various uses. A wide variety of control strategies can be considered, including holding, deadheading, expressing, and short-turning strategies. These range in complexity and in data requirements, ranging from the simplest headway-based holding strategy which requires only the predicted arrival time of the trailing vehicle, to short-turning strategies which ideally would require real-time data on passenger loads.

The benefits of a decision tool include not only the implementation of control strategies, but also the archived data it produces, which can be used to improve performance analysis and service planning. When control strategies are applied in an ad-hoc way through radio or in-person communication, typically no records are made of the actual control decision, and so users of archived data must infer whether or not a control action was taken. Archived data from an automated decision tool allows much more precise knowledge of whether or not a control action was recommended, which then allows for more accurate tracking of on-time departures, dwell times, and run times.

As described in this chapter, a variety of data sources, decision algorithms, and user-interface options exist, but all must fit into a basic structure for downloading the data, applying an algorithm, outputting the suggested control actions to users, and archiving the control actions. Typically a "prediction interpreter" component will be needed to convert customer-facing data sources, which focus on predicted arrival times in the near future, into a useful format for decision tools, which often require a recent history of departure times.

Chapter 4

Experiment

This chapter describes the results of an experiment performed on the MBTA's Route 1 bus service, using the mobile app described in Chapter 3. During the experiment, irregular terminal departures were observed, which prompted further investigation of causes of and potential solutions to these irregular departures. These are also explored in this chapter. Recommendations are made aimed at improving regularity of departures and enabling more precise control of departure times.

4.1 Description of experiment

The experiment was conducted from 2:30 PM to 6:30 PM each day from Monday, September 8, 2014 through Friday, September 12, 2014. In this section, we describe the experiment, including Route 1, the strategy used, and the personnel who implemented the strategy.

4.1.1 Route 1

Route 1 is a crosstown route running from Harvard Square in Cambridge to Dudley Square in Roxbury, mostly along Massachusetts Avenue, a major street that is often congested. The route connects a variety of employment centers, residential neighborhoods, and commercial areas. The published map is shown in Figure 4-1.

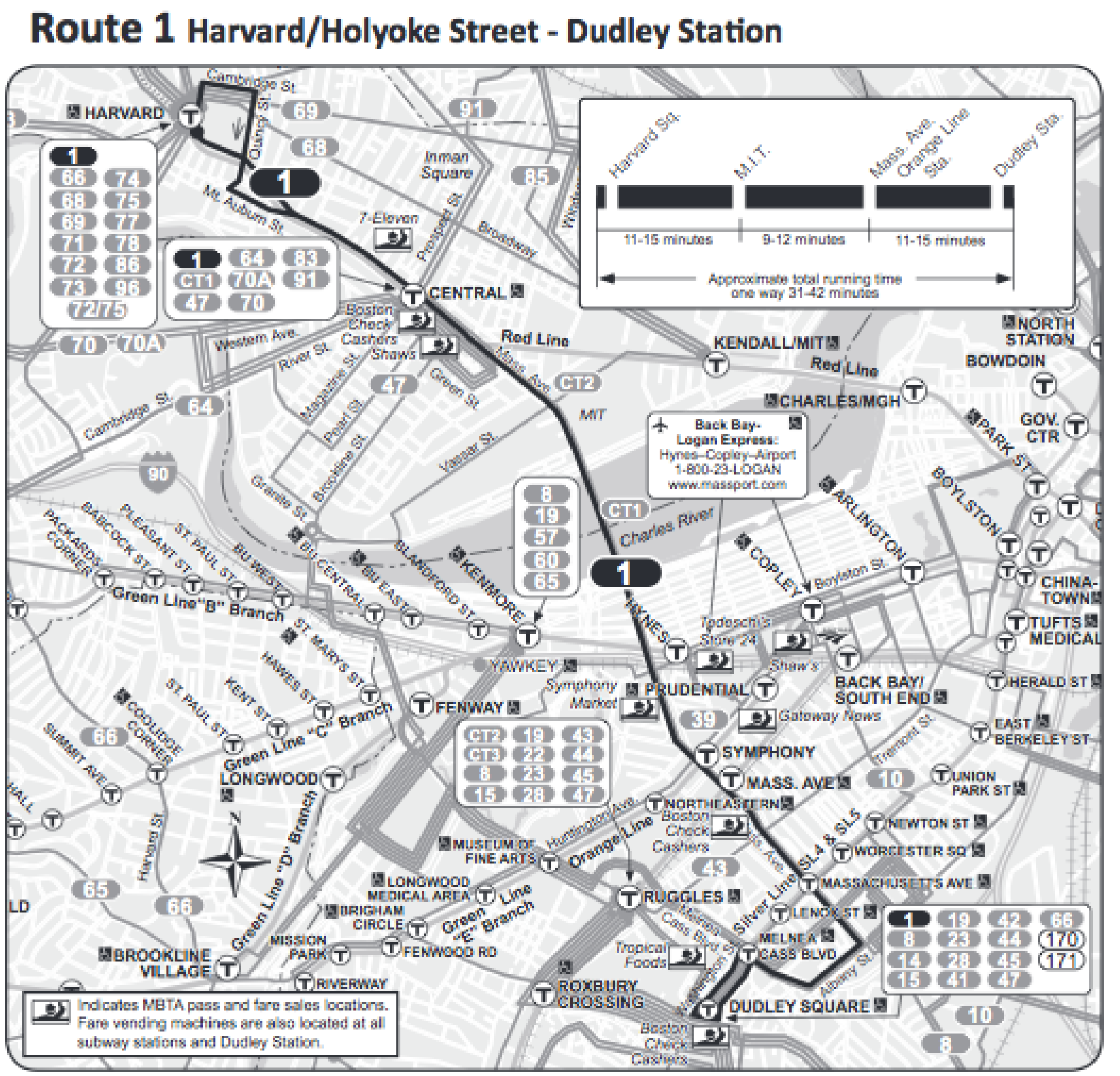

Figure 4-1: Published map of MBTA Route 1

4.1.2 Strategy

The strategy used is a variant of Turnquist's prefol strategy, which was described in Section 3.2.1. In this version of the strategy, when a bus arrives at a terminal, a departure time is selected based on the departure time of the previous vehicle to leave the terminal, and the predicted departure time of the trailing vehicle. The ideal departure time, according to the strategy, is the average of these two times. We constrain it to be no earlier than the scheduled departure time. This simplifies calculations and ensures that downstream passengers relying on NextBus predictions (which may be schedule-based) will not miss a bus due to holding.

4.1.3 Personnel

Implementation of the strategy was accomplished with two people stationed at each terminal: One inspector and one researcher. Inspectors, in MBTA Bus Operations, are supervisory personnel who manage various aspects of bus service, including managing garage pull-outs and pull-ins, responding to bus breakdowns or accidents, assisting passengers, adjusting service in response to delays, and many other field responsibilities. Bus operators are required by their contract to follow any special instructions given by inspectors. The inspectors for this experiment signed up for a special overtime detail covering the period of the experiment; on days when no inspector had signed up, either an inspector was pulled from mobile "radio car" duty or a higher-ranking official from the Southampton garage covered the detail. The role of the inspector was to read the assigned departure time from the app, instruct bus operators to depart at the assigned time, and observe the operators to ensure compliance.

In addition to the inspector, one researcher was present, either a student from MIT or a member of the MBTA's research or IT staff. The role of the researcher was to bring the mobile device (one phone and one handheld computer were used), start up the app, and explain the use of the app to the inspector. After explaining the app and answering any questions, the researcher was not needed until the end of the shift, when they would return to pick up the device. However, in most cases the researcher would remain and observe the experiment for at least one or two hours.

4.2 AVL data analysis for three typical weeks

The weekdays of the three weeks following the experiment, running from September 15 to October 3, 2014, were used as a "control", both as a baseline for comparison and for insight into the typical causes of departure delays. These weeks were selected as they were during the academic year and contained no holidays, and were thus comparable to the experiment week of September 8 - 12. We first consider the effect of scheduled cycle times on performance, then examine how a vehicle's headway departing the terminal propagates along the route, and finally compare available recovery time to schedule deviation to determine how many early or late departures can be attributed to operator behavior.

4.2.1 Recovery time and half-cycle time

Available and actual recovery time

Tables 4.1 and 4.2 describe the recovery times at the terminals. "Scheduled recovery time" refers to the scheduled time between an arrival of a vehicle at Harvard and the next departure of the same vehicle within the same block. "Available recovery time" refers to the difference between the arrival time and the next scheduled departure time for that vehicle, which may be negative if the vehicle arrived after its next scheduled departure time. "Actual recovery time" is the actual time that elapsed between a vehicle's arrival and its next departure.

Most notable is that available recovery times at Harvard during the observation period were insufficient during the Midday and PM Peak periods, when the median available recovery times were very low. This could be caused by either insufficient cycle time in the schedule, or late departures from Dudley. To determine the cause, we examined the half-cycle times in each direction.

Table 4.1: Recovery Times at Harvard

Time Period	Trips Observed	Median Recovery Time (mins)			Median Lateness of Departure
		Scheduled	Available	Actual	
AM Peak	182	11.0	7.8	8.2	0.2
Midday Base	275	9.0	1.1	2.7	1.6
Midday School	147	10.0	1.7	3.5	2.2
PM Peak	230	9.0	-0.3	2.9	3.0
Evening	316	10.0	6.0	6.2	0.3
Other (Late/ Early AM)	421	11.0	5.8	6.4	0.4

Table 4.2: Recovery Times at Dudley

Time Period	Trips Observed	Median Recovery Time (mins)			Median Lateness of Departure
		Scheduled	Available	Actual	
AM Peak	182	9.0	8.4	10.3	1.3
Midday Base	279	12.0	6.8	9.2	1.7
Midday School	178	11.0	5.4	6.9	1.5
PM Peak	245	12.0	4.1	7.4	3.2
Evening	321	10.0	8.2	10.0	1.1
Other (Late/ Early AM)	473	8.0	4.3	4.7	1.1

Half-cycle time analysis

To show that insufficient half-cycle time is the cause of the lack of recovery time, we compare the scheduled half-cycle times with the distribution of observed run times. Tables 4.3 and 4.4 show the scheduled half-cycle times in each direction and the median, 90th, and 95th percentile of run times. The 95th percentile of run time is commonly used to set scheduled cycle times, with the goal that 95% of trips will arrive before the next scheduled departure. At some agencies the 90th percentile is used.

The 95th percentiles of run times for the three-week baseline period exceed the scheduled half-cycle time for trips from Dudley to Harvard in the Midday and PM Peak periods. In the PM Peak, there is an offsetting amount of extra time available in the Harvard-to-Dudley direction that could be shifted. However, in the Midday School and Midday Base periods the half-cycle time is too short in both directions, and therefore does not have such a simple solution.

Table 4.3: Half-cycle time analysis - Trips from Harvard to Dudley

Time Period	Median Run Time	90th Pctile Run Time	95th Pctile Run Time	Median Sched. Half-Cycle Time	Extra Half-Cycle Time vs. 95th pctile
AM Peak	40.6	47.6	50.7	49.0	-1.7
Midday Base	41.3	50.2	53.2	54.0	0.8
Midday School	49.1	60.5	63.7	58.0	-5.7
PM Peak	46.2	54.6	56.9	60.0	3.1
Evening	35.1	40.8	42.6	46.0	3.4
Other (Late/ Early AM)	27.6	34.5	35.7	33.5	-2.2

Table 4.4: Half-cycle time analysis - Trips from Dudley to Harvard

Time Period	Median Run Time	90th Pctile Run Time	95th Pctile Run Time	Median Sched. Half-Cycle Time	Extra Half-Cycle Time vs. 95th pctile
AM Peak	37.5	43.4	47.1	44.5	-2.6
Midday Base	42.0	53.2	55.4	45.0	-10.4
Midday School	40.7	51.6	55.8	46.0	-9.8
PM Peak	42.6	52.3	55.3	52.0	-3.3
Evening	32.5	39.2	41.9	42.0	0.1
Other (Late/ Early AM)	25.2	32.3	35.9	32.0	-3.9

4.2.2 Headway variance along the route

In this section, we use a framework developed by Cham (2006) to examine the effect of departure headway on downstream performance. In Appendix C, Tables A.1 and A.2 show how headways vary along the route, using the ratio of actual headway to scheduled headway at the terminals and at Hynes as a major midpoint stop. Consistent with Cham's findings and those of other authors, the coefficient of variation of headways increases as buses travel along the route. Also, buses that begin their trips with headways close to the scheduled headway have a lower coefficient of variation throughout the trip, and buses that depart with a headway ratio between 0.8 and 1.2 generally maintain these headway ratios throughout the entire trip.

4.2.3 Deviation from scheduled departure time

Figures 4-2 and 4-3 show the distribution of deviations from scheduled departure times at Harvard and Dudley, grouped by the available recovery time. Shades of green are used for trips with 2 or more minutes of recovery time available, and red and orange for trips with insufficient or negative recovery time. Cham used this to estimate the minimum amount of time needed to turn around at Dudley on the Silver Line at approximately 2-3 minutes. We find a similar result, observing that recovery times of 2-4 minutes were sufficient for most operators to depart on-time at both terminals.

By graphing the schedule deviations in groups by amount of recovery time, we can extend Cham's analysis. Most notable is that on the Harvard graph, on-time performance improved with additional recovery time, increasing up to the groups that had at least two minutes of recovery time, which all had very similar distributions tightly clustered around the scheduled time. At Dudley, on the other hand, buses with two to four minutes of recovery time had the best performance, but buses with *more* than four minutes of recovery time had *worse* performance. This may be because operators are more likely to take breaks at Dudley than at Harvard, and are more likely to take those breaks when they have more recovery time available.

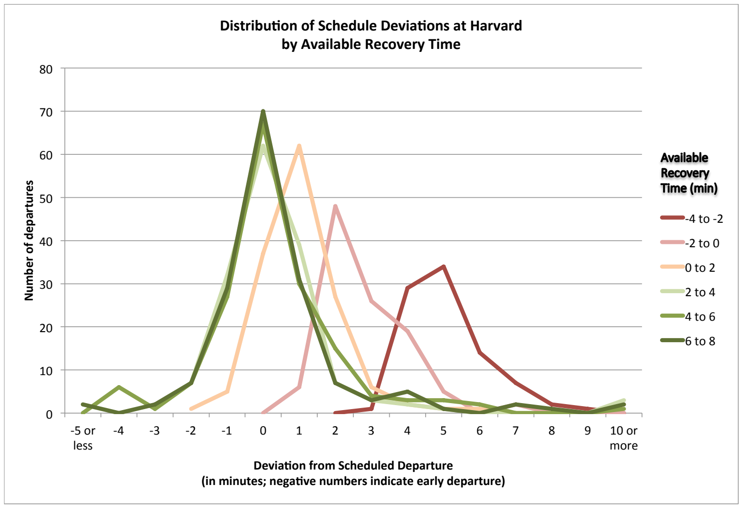

Figure 4-2: Schedule Deviations at Harvard by Available Recovery Time

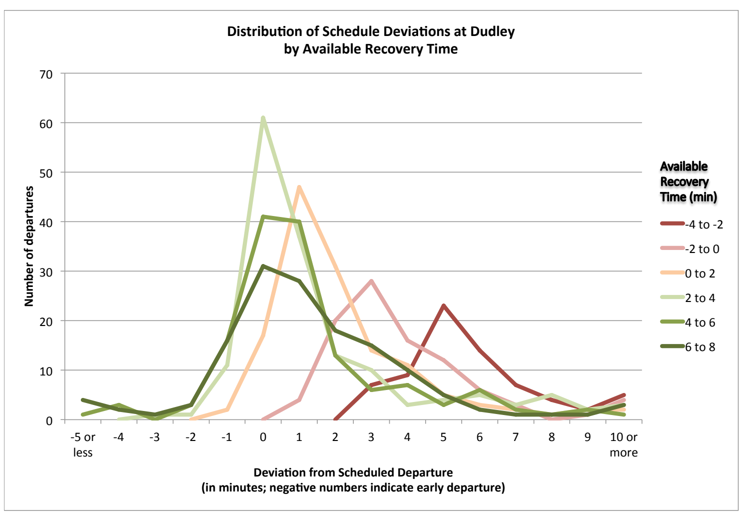

Figure 4-3: Schedule Deviations at Dudley by Available Recovery Time

4.3 Summary of holding instructions given and deviation from instructions

In this section, we examine the range of instructions that were given by the app to supervisors, and the compliance of the supervisors with these instructions. The data from Friday September 12 were excluded from this analysis because there were severe delays, only one inspector was available (stationed at Harvard), and alternative strategies were attempted.

Of the 184 trips made during the experiment:

- 15 were instructed to depart on schedule
- 31 were instructed to depart "ASAP"
- 121 were given holding instructions, per the prefol strategy
- 17 have unknown status, due to missing data

4.3.1 Holding instructions

Here we summarize the holding instructions that were given during the experiment, to illustrate the impacts of the strategy on regular bus operations. Appendix B shows the distribution of suggested layover times; that is, the difference between the arrival time of a bus and its suggested departure. Recommended layover times varied widely, from 0 minutes up to 25 minutes. Figure 4-4 shows the distribution of lateness of suggested departure times vs. the schedule. This is important as holding buses too far past their scheduled departure times may eventually incur additional costs in overtime pay for operators, if an operator is unable to make up the delayed time on later trips. Buses given the "ASAP" instruction are excluded because they are not being held for any extra time. The figure shows that the majority of suggested departure times were later than the scheduled time by 8 minutes or less.

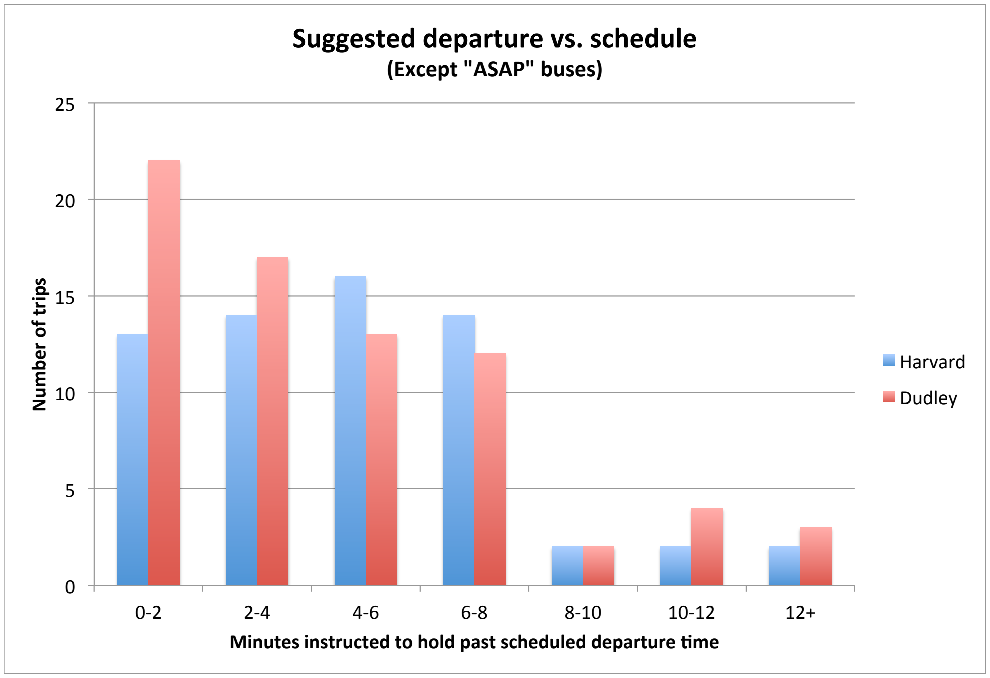

Figure 4-4: Suggested Departure vs. Scheduled Departure at Each Terminal

4.3.2 Deviations from suggested departure times

In this section, we examine deviations from the suggested departure times given by the app. For this purpose, we separate departures into those made with instructions to depart "ASAP" and all other types of departures.

"ASAP" departures

We first examine those trips for which the bus arrived later than its scheduled departure time and the app advised a departure "ASAP". For these trips, operators and supervisors both had a strong incentive to begin the next trip as soon as possible. Thus, the realized recovery times in these scenarios will shed light on how much recovery time is needed at each terminal. The distribution of recovery times is shown in Figure 4-5. Of the 17 such trips at Harvard, the recovery time had a mean of 2.3 minutes and a standard deviation of 0.7 minutes. On the other hand, at Dudley, the 14 departures in this group had a mean of 4.6

minutes and a standard deviation of 2.6 minutes of recovery time, indicating a significantly longer and more variable turnaround process.

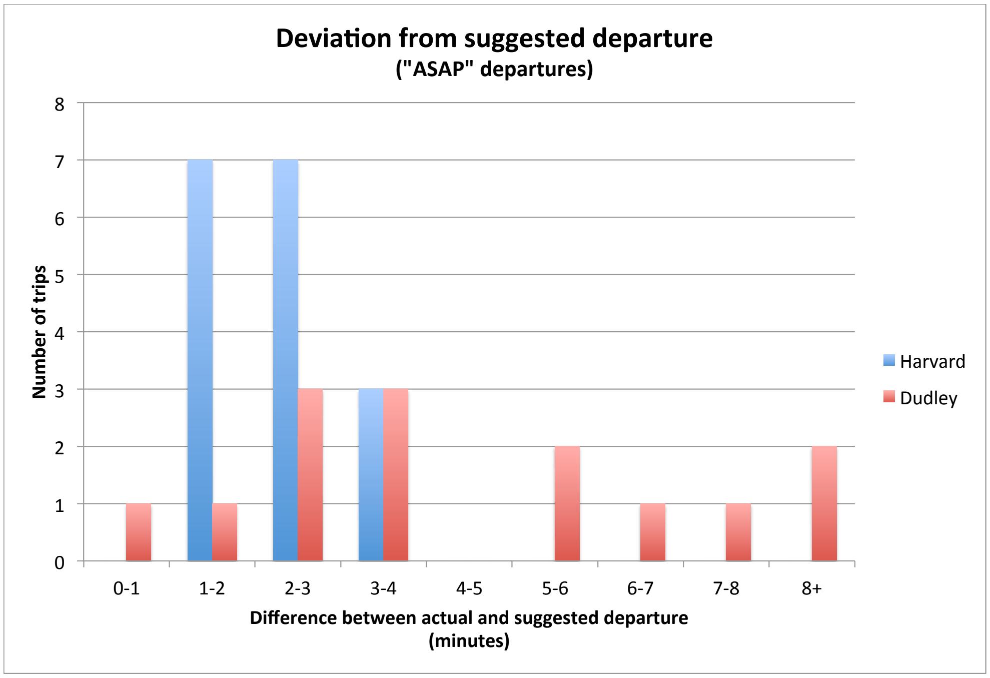

Figure 4-5: Deviation from suggested departure with “ASAP” instructions

Other departures

Figure 4-6 shows deviations from the suggested departure times among trips other than the “ASAP” trips described above (including both “on-schedule” departures and those that were held). Although in theory the inspectors stationed at the terminals should have been able to control the departure times, it is clear that this control was frequently ineffective. We draw particular attention to early departures, which can only indicate poor performance by the supervisor at the terminal. Operators may share the blame if they knew the suggested time and departed early anyway, but since all instructions to operators came through direct communication from supervisors, departure-time control during the experiment must be primarily driven by supervisor diligence.

To see if there was any pattern of poor performance by specific supervisors, we examined the number of departures that occurred at least 2 minutes early by date and terminal in Table 4.5. This shows that 8 of the 10 total departures that were made more than 3 minutes early were made under the supervision of just 3 of the 12 supervisors: The supervisor at Dudley on 9/10, and those at both terminals on 9/8.

The supervisor at Dudley on 9/10 alone was responsible for four of these extremely early departures, suggesting that something went seriously wrong with the implementation of the strategy on that date. Discussion with MBTA staff revealed that there was some delay in delivering the handheld device to be used with the app to Dudley that day. However, this alone cannot explain the deviations, since there were major deviations throughout the shift, not just at the beginning. The same inspector was on duty the following day at Dudley, and had no such early departures, so it may be that specific conditions on that day caused the inspector to be unable to implement the strategy correctly. Ensuring effective supervision will be key to any future implementation of this strategy, and we will recommend that inspectors be given strict instructions and informed about what data is being gathered on the performance of the strategy, in an effort to encourage compliance.

Table 4.5: Early departures by date and terminal

Date	Departure Terminal	# of Departures 3+ Minutes Early
9/8	Harvard	2
9/8	Dudley	2
9/9	Harvard	0
9/9	Dudley	0
9/10	Harvard	0
9/10	Dudley	4
9/11	Harvard	1
9/11	Dudley	1

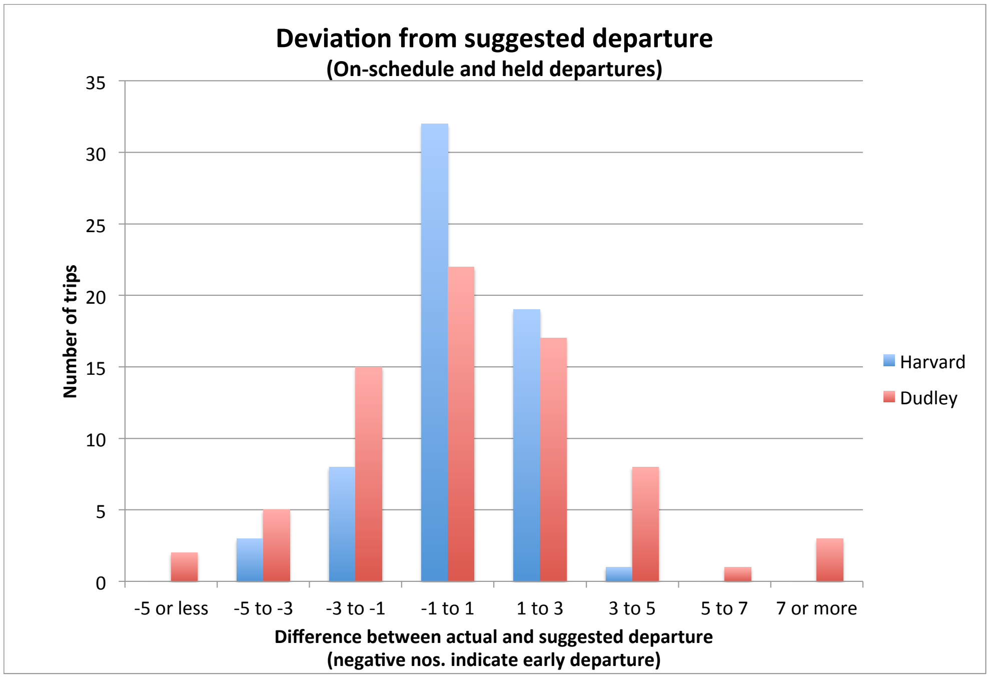

Figure 4-6: Deviation from suggested departure at each terminal

4.4 Impact of the experiment on headway regularity

4.4.1 Coefficient of variation of headway

To determine the impact of the experiment, we first examine the coefficient of variation of headways (the standard deviation divided by the mean) during the experiment, and compare it to the baseline period of the three weeks following the experiment, excluding Fridays. The results are shown in Figure 4-7 for trips departing Harvard, and Figure 4-8 for trips departing Dudley. The results differ strongly between the two terminals.

For trips departing Harvard, the coefficients of variation were lower during the experiment period than the baseline period, at all timepoints except for the second-to-last timepoint during the PM Peak period. The magnitude of the improvement in the midday period was strong all along the route, while in the PM Peak the effect decreased as buses got closer to Dudley. This is likely a result of higher variation due to traffic during the PM Peak period

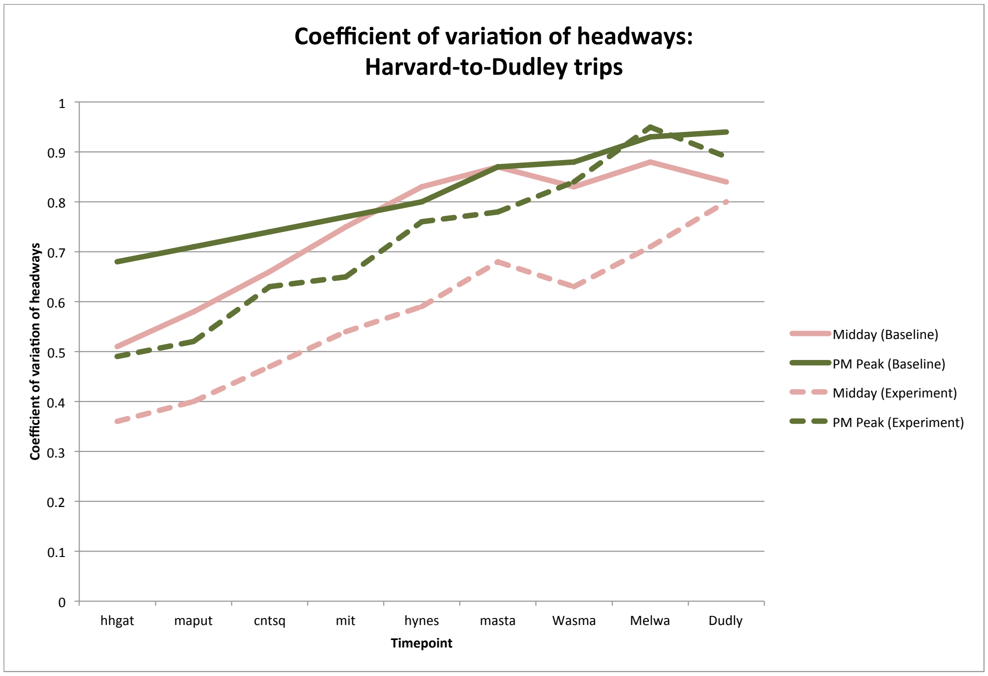

Figure 4-7: C.O.V. of headways - Harvard to Dudley

making the effects of holding at terminals less persistent throughout the route.

For trips departing Dudley, the coefficients of variation during the experiment were generally similar to or worse than those during the baseline period. Factors contributing to the poor performance at Dudley will be explored in the next section.

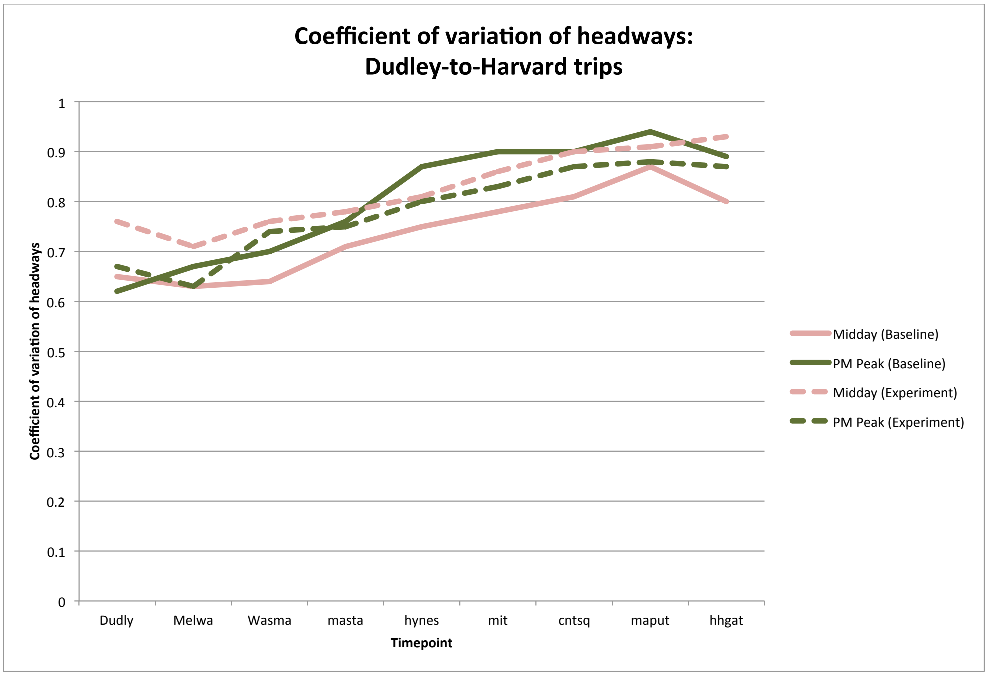

Figure 4-8: C.O.V. of headways - Dudley to Harvard

4.4.2 Average passenger wait time

Average wait time experienced by passengers at a stop is calculated using the formula $W = \frac{1}{2} \cdot \mu_h \cdot (1 + (\frac{\sigma_h}{\mu_h})^2)$, where μ_h is the average headway between vehicles at that stop, and σ_h is the standard deviation of the headways of vehicles at that stop. Figures 4-9 and 4-10 show the average passenger wait time along the route in each direction. The weighted average wait time is calculated using the number of boardings at each stop during the relevant time periods. It is assumed that passengers at non-timepoint stops experienced the same headway as at the nearest timepoint.

The results are broadly similar to the coefficients of variation of headways, with an overall average improvement of approximately one minute in passenger wait time for trips originating at Harvard, and no improvement for trips originating at Dudley. In the Harvard results, it is interesting to note that the improvement of average passenger wait time is higher in the middle of the route than at the beginning. Even though the coefficient of variation

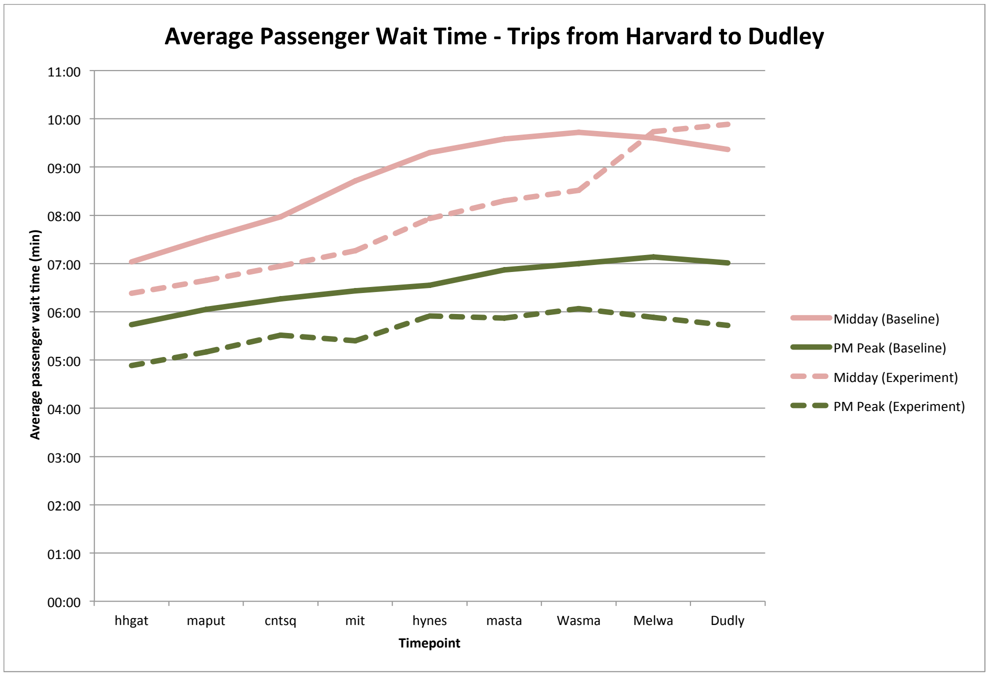

Figure 4-9: Average passenger wait time by timepoint - Harvard to Dudley

did not improve as much at the midpoint, the higher magnitude of wait times meant that this is where the largest absolute improvement in wait time was found.

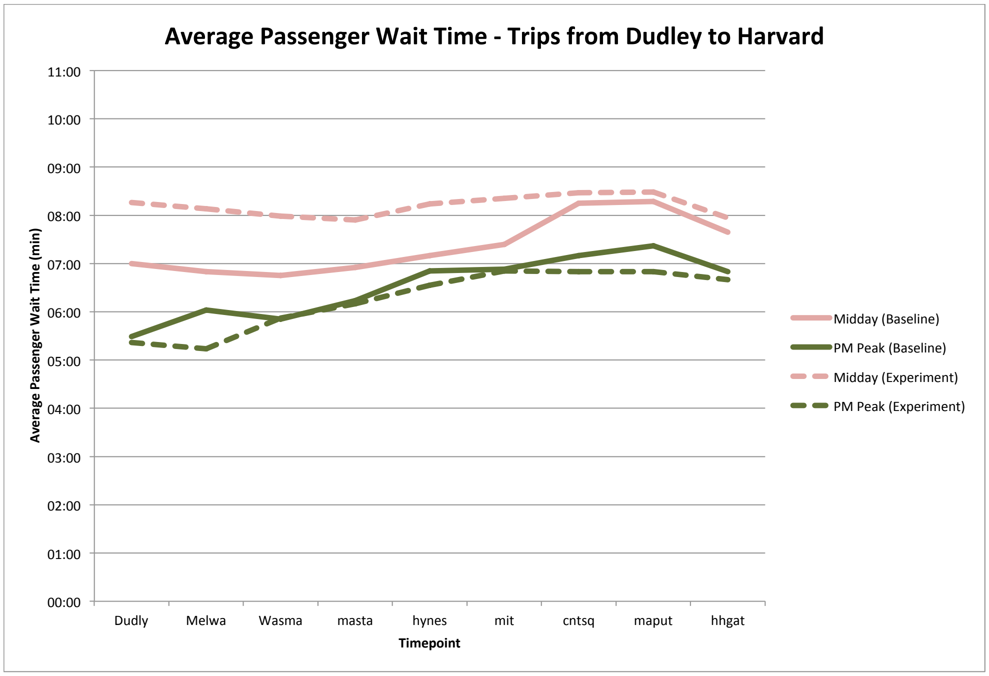

Figure 4-10: Average passenger wait time by timepoint - Dudley to Harvard

4.5 Factors contributing to unreliability at Dudley

It is clear, based on the deviations from the app suggestions, as well as the high variation in departure headways, that headway control at Dudley was not effective. Three factors stood out as contributors to large variation in departure times at Dudley: Operator discipline and supervision, the layout of Dudley station, and the long boarding times of passengers at Dudley.

4.5.1 Terminal departure discipline

The most important aspect of operations at Dudley leading to irregular departure times is operator behavior, specifically, operators returning late from breaks or departing early. Operators frequently take breaks while laying over at Dudley, due in part to easy access to a bathroom. If the operator does not return to the departure area before the next departure time, a late departure results.

Terminal departure discipline is a core issue at all transit agencies. Using real-time and archived AVL data, it can be addressed through improved scheduling, operator training and discipline, and the use of real-time displays at terminals. In this section, we examine these methods of improving departure discipline, with a focus on methods that have been used successfully at other transit agencies.

Scheduled recovery time

One reason breaks at Dudley are more popular is the scheduled cycle time mentioned previously; because there is not enough time scheduled for trips from Dudley to Harvard, operators rarely have any time available to take a break at Harvard, and thus are much more likely to take breaks at Dudley. Insufficient cycle time is an obvious contributor to poor on-time performance at terminals, as an operator who arrives at a terminal later than his next scheduled departure time cannot possibly depart on time.

In the context of our holding strategy, the lack of sufficient recovery time may have had a positive impact on implementation at Harvard. Operators who arrive at Harvard with little or no recovery time typically do not leave their vehicles, making it easier for an inspector to control departure time. Adjusting the scheduled cycle time to increase recovery time at Harvard might actually have the effect of reducing the effectiveness of a holding strategy by creating more opportunities for operators to leave the bus stop area. If a holding strategy were to be implemented on a permanent basis during peak hours, it would be important to carefully consider the appropriate amount of recovery time at each terminal.

Supervisor effectiveness

The ineffectiveness of supervisors during the experiment contributed to poor departure-time adherence. Section 4.3.2 discussed in particular the incidence of very early departures, but in general, the frequency of both early and late departures relative to the suggested times indicates that supervisors did not strictly adhere to the strategy. This implies a need for an improvement in the culture of supervisors, starting with training and including continuing monitoring of performance. Supervisors must be aware of the metrics used to evaluate performance, and understand the connection between their control actions and the customer experience.

One example of an agency culture change comes from the Chicago Transit Authority, which undertook several agency-wide initiatives in 2007-2008 under President Ron Huberman. The CTA focused on management-level interventions, with weekly meetings between upper management and garage managers. For these meetings, a report was prepared with a table of "Big Gap" metrics for each route and "Key Levers", meaning factors impacting performance such as early or late departures, accidents, absences, etc. This focus on accountability at the garage management and supervisor level, not just the operator level, is important to creating a culture of departure-time discipline. The CTA was able to reduce the rate of late departures from 15.3% to 10.6%, and early departures from 5.0% to 2.7%.

Another important part of supervisor culture is the use of technology. Currently, the use of technology by supervisors varies widely. Some actively use the handheld devices provided by the MBTA to observe bus locations and statuses on the TransitMaster screen, in support of their role. Others ignore the technology available and make interventions based only on information available via radio or other methods. Explanations given by supervisors include the difficulty of use of the handheld devices, which are large and run software designed for desktop computers. The use of technology in support of the supervisory role must be emphasized in training, and more user-friendly mobile technologies developed, oriented towards use in the field rather than in the control center.

Operator training and performance monitoring

Creating a culture of on-time terminal departures among operators is a challenge for all transit agencies. The influence of the agency on operator culture includes: 1) Training 2) Monitoring performance and 3) Re-training. Automatically collected data provide an opportunity to analyze and influence operator behavior. We spoke with staff from RTD in Denver and Metro Transit in Minneapolis to find current industry best practices in improving on-time performance. They focused on the use of data to monitor and re-train operators, but the same data and concepts should also be emphasized during training to create an expectation of on-time performance from the start of an operator's career.

Both RTD and Metro Transit use AVL data to create operator on-time performance reports. They emphasize the importance of limiting analysis of late departures to cases in which the operator had sufficient recovery time to depart on-time. RTD began an initiative in May 2014 to use performance reports to initiate conversations between division managers and bus operators about on-time departures, which they defined as 15 seconds early to 1 minute late. They compared operators who had an intervention from a manager to those who did not, and found a statistically significant improvement of 15% in on-time departures vs. a control group improvement of 6%.

Metro Transit follows two strategies to use AVL data to improve departure-time discipline. One is a similar report to that used by RTD, showing early or late departures for managers to discuss with operators after the fact. A second strategy is used in real time: Using a setting in the TransitMaster AVL system, control-center staff are able to see pop-up alerts when a bus departs early from a terminal. They can then communicate directly to the bus operator, providing immediate feedback and a clear reminder that early departures are unacceptable.

Two examples of operator departure-time analysis are given below. Table 4.6 shows the percentage of departures made 1 or more minutes early by each operator. Table 4.7 shows departures made 3 or more minutes late in cases when there was at least five minutes of available recovery time. This sample output uses the baseline time period of the three weeks following our experiment, and all departures from terminals made by operators from the Cabot garage who made at least 20 trips during the study period are included.

Table 4.6: Early terminal departures by operators from Cabot Garage, Sep. 15 - Oct. 3, 2014

Operator	Total Trips	Departures >1 min Early	% Early Departures
A	96	56	58.3%
B	145	80	55.2%
C	138	59	42.8%
D	77	32	41.6%
E	142	59	41.5%
95th percentile			20.8%
Median			3.9%

For comparison purposes, the median values and 95th percentiles of the early departure and late departure statistics are included, and show that the values seen in this table are abnormal, and that a small number of operators are causing a large fraction of early or late departures. Managers should have access to this information so that they can speak to these operators about their on-time performance. Operator behavior must be the first area to improve, because if operators do not comply with schedules and instructions, the other potential improvements listed below will be useless.

Table 4.7: Late terminal departures by operators from Cabot Garage, Sep. 15 - Oct. 3, 2014

Operator	Total Trips	Departures >3 min Late	% Late Departures
F	136	28	20.6%
G	94	17	18.1%
H	193	31	16.1%
I	90	14	15.6%
J	137	19	13.9%
95th percentile			11.8%
Median			3.2%

Real-time displays

Encouraging on-time performance can be achieved through visual or auditory cues at the terminal. One strategy that is used on the MBTA heavy rail system is the "ring-off" bell, which indicates to both train operators and customers that the train is about to depart, putting pressure on the train operator to depart quickly. At a station like Dudley, where many bus routes pass through, a bell might not be sufficient to indicate a specific route's imminent departure. An alternative would be to use the departure boards that list the next scheduled departure time. LED boards capable of displaying text have recently been installed at several bus berths at Dudley, but are not currently in use. These should be used to display the next departure time of each route at its actual berth which could be a better way of informing both operators and customers of the next departure time, and encouraging operator compliance. In addition, by adjusting the departure times displayed, these signs could provide a way to implement terminal holding strategies in the future. They would allow dispatchers or an automated system to communicate special departure instructions to bus operators easily, without the need for a supervisor on-site.

4.5.2 Boarding times

During the experiment, we observed unusually long boarding times at Dudley. One reason for this was a large number of passengers adding value to their stored-value CharlieCards as they boarded the bus. The bus fare is \$1.60 with a CharlieCard vs. \$2.10 when paid directly with cash, incentivizing customers to use the stored-value option. The Roxbury-Dorchester-Mattapan Transit Needs Study (2012) found that in lower-income communities such as those along the bus routes that feed into Dudley, passengers add value to their cards more frequently and have less access to off-board methods of adding value. The study recommended increased availability of CharlieCard vending machines in the neighborhood as well as policy changes such as a change in the fare structure or a minimum level of value to add on a bus. There is a CharlieCard vending machine at Dudley station, but it is not well-utilized and is located far from the Route 1 berth.

Another reason for the long boarding times was a higher number of passengers in wheelchairs. The process of deploying the ramp, boarding, strapping the wheelchair in, and resetting the ramp took several minutes, during which no other passengers could board.

To mitigate the problem of long boarding times, one possible strategy is the use of handheld CharlieCard validators. The MBTA already uses such validators on the Green Line, and is in the process of purchasing additional validators for use by Bus Operations. The inspector assigned to the Silver Line area at Dudley can see the Route 1 and other berths from his typical post, and intervene when a bus is experiencing a slow boarding process. We will recommend that inspectors at Dudley be provided with a handheld validator to speed up the boarding process whenever they observe a long queue of passengers waiting to board.

4.5.3 Dudley station layout and operations

This section describes factors relating to the layout of Dudley Station and typical operations at the station that contribute to unreliable departures.

Layout of Dudley Station

The layout of Dudley Station is shown in Figure 4-11, and in Figure 4-12 we show the area used by Route 1 buses. There are two lanes used by four different bus routes, one for pulling up to the curb to load and unload passengers, and the other for laying over or leaving buses for a change of operator (a “swing-off”). Passengers are unloaded at or near Berth 18, and buses on Route 1 typically then pull into the layover area, as far up to the front as possible, to allow for more buses to pull in behind them. Usually buses occupy at least some part of the layover space, and as many as four buses have been observed occupying this area at one time. Route 1 is the only route with berths in this section that terminates at Dudley, but occasionally buses from other routes such as Route 66 may also use this layover lane.

Conflicts between routes

Delays were observed in situations where conflicts existed between the different routes and parked buses. For example, a bus parked at the back of the layover area and needing to pull in to pick up passengers for departure could be in conflict with a bus picking up passengers in Berth 19, especially if there are more buses parked in front of it. Layover space is not necessarily used in a "first-in, first-out" manner because buses may arrive out of order or may have scheduled layovers that overlap due to interlining or crew changes.

Another problem caused by the layout of the station is the difficulty in loading passengers early, before the scheduled departure time. Operators at Harvard were frequently observed allowing passengers to board a few minutes before the scheduled departure time, allowing them to depart immediately once the scheduled departure time is reached. At Dudley, however, pulling up to the berth early can lead to blocking buses from other routes, in situations where the layover space next to Berth 19 or Berth 20 is occupied. Figure 4-13 shows an example of a scenario where a Route 8 bus is blocked by a Route 1 bus loading at Berth 20 and another Route 1 bus laying over.

Reversing difficulty and "loop-around" maneuver

Reversing a bus is a difficult maneuver even in the best of circumstances, and in the crowded conditions at Dudley, it is even more difficult. During a follow-up visit to Dudley on November 10, 2014, operations on Route 1 were observed between 2:30 PM and 5:30 PM, with arrival times, operator break times, passenger loading times, and departure times recorded, along with other notes.

One key observation was that buses parked far up in the layover area either had to reverse in order to pull up to the curb at Berth 20, or had the option of pulling out of the station and looping around via Warren Street and Washington Street to re-enter the station. Operators looping around in this way typically left 2-3 minutes prior to the scheduled departure time, to compensate for the added time. After pulling up to the berth and loading passengers, all operators then departed immediately, regardless of the scheduled departure time. Looping around is undesirable because of the uncertainty it adds to departure times, and the changes we propose for Route 1 layover procedures at Dudley should eliminate this as an issue.

Proposed procedure

We propose a revised use of layover and passenger loading areas at Dudley. It entails using the Route 1 berth as a space to lay over, and blocking off a part of the layover area to allow other buses to pass the Route 1 berth. The primary goal of this procedure is to allow bus operators to pull up and begin boarding in advance of the departure time.

1. Upon arrival at Dudley, drop off passengers at or behind Berth 18.

2. If the current bus is the next bus to depart from Dudley:

   (a) Pull up to park and lay over at Berth 20.

   (b) If more than two minutes of recovery time is available before the next scheduled departure, the driver may leave the bus for a regular break.

   (c) Two minutes before the scheduled departure time, the driver should return to the bus and begin to allow passengers to board.

3. If the current bus is not the next bus to depart from Dudley:

   (a) Pull up as far forward as possible within the layover area, without blocking the area marked as "No Layovers" in Figure 4-14. (This area should be marked with a sign or paint).

   (b) The driver should return to the bus with enough time to pull up to the berth and begin boarding passengers at least two minutes prior to the scheduled departure time.

In addition to allowing for early boarding, this procedure should eliminate some of the conflicts with other bus routes and eliminate the requirement for buses to occasionally loop around local streets. The cost would be only the reduced layover space shown in Figure 4-14. This still leaves sufficient space for four buses to lay over, which is equal to the maximum number of buses observed utilizing the space at any time during the course of this research.

This strategy should be combined with the use of the LED sign to display the next departure time as described above. There are two reasons for this: The first is that passengers will, at least initially, be confused by a bus laying over at its usual departure berth, and the LED sign will clearly communicate the next departure time. The second reason is that it is not always clear to a bus operator whether or not their bus is the next in line for departure, for example if the leading bus has fallen behind or pulled out of the garage late. The LED sign will provide an easy way for the operator to see if their departure is the next one, and thus, if they should pull up to the berth or park in the layover area.

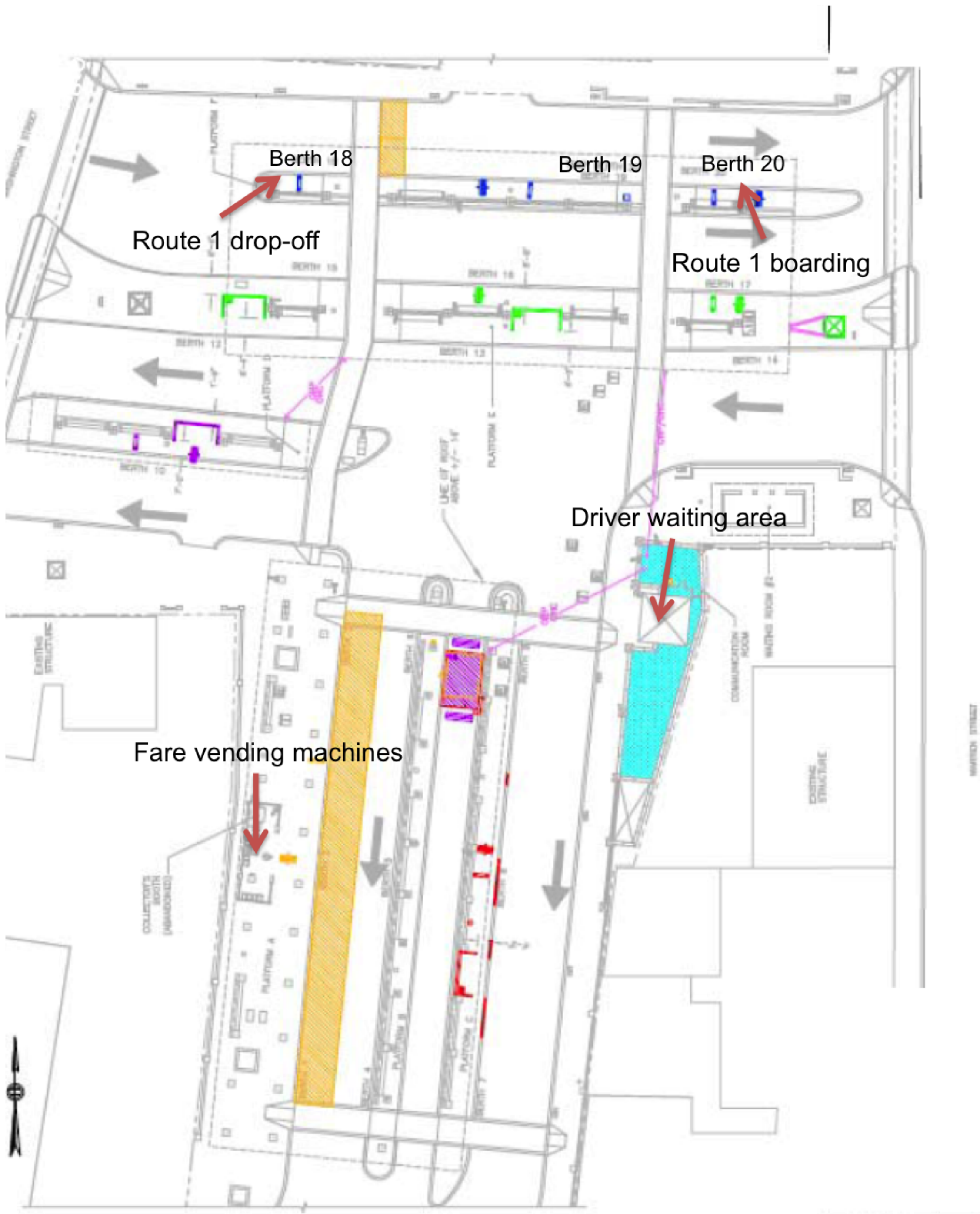

Figure 4-11: Diagram of Dudley Station

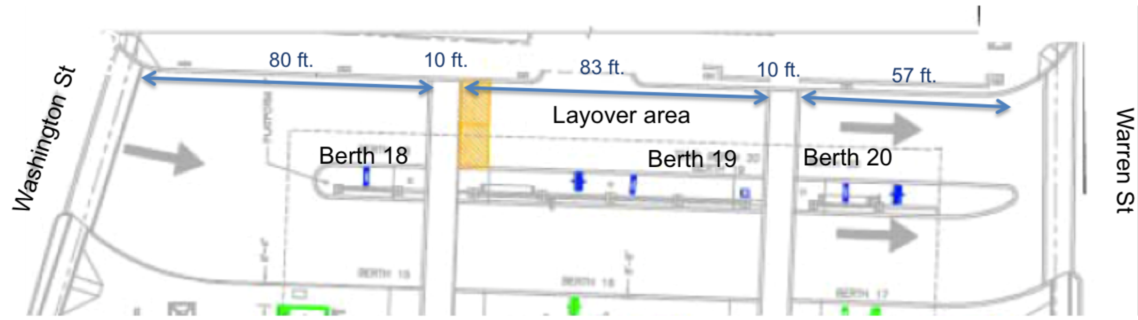

Figure 4-12: Route 1 loading and layover area at Dudley Station

Figure 4-13: Blocking scenario

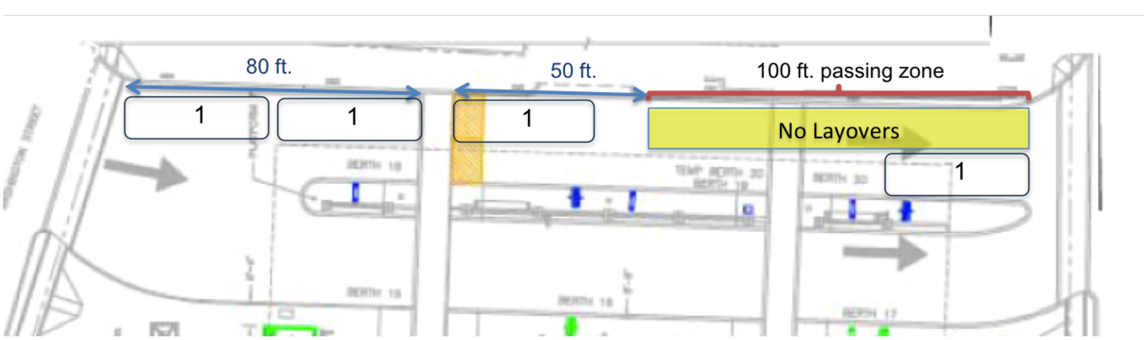

Figure 4-14: Proposed layout

4.6 Deadheading and expressing strategies

In addition to the holding strategy used in this experiment, several control strategies exist that allow buses to reduce big gaps by skipping some scheduled stops. During the experiment and the earlier test run in April 2014, inspectors at Harvard Square were observed implementing different strategies, including one used on the Friday of the experiment that was clearly inferior to the regular route. In order to allow for the best use of real-time data, a standardized strategy or set of strategies must be developed. In this section, we examine the deadheading, expressing, and short-turning strategies available for Route 1 at Harvard Square. We pay particular attention to strategies that can be implemented with little or no visibility to customers, that is, without denying boardings or directly bypassing stops.

4.6.1 Types of strategies

A variety of control strategies are used by transit agencies to close a large gap ahead of a vehicle. These include short-turning, in which a bus discharges passengers and turns back without reaching the terminal, deadheading, in which a bus departs the terminal without picking up passengers and skips some number of stops, and expressing, in which a bus first picks up passengers, then skips some stops. Several such strategies are available at Harvard due to the dense street grid near the terminal stop, and the circuitous route normally taken by Route 1.

Figure 4-15 shows four possible routes for the turnaround at Harvard: (1) is the standard route, (2) is the version used by the inspector at Harvard on the Friday of the experiment, (3) is a short-turn down Bow Street that was observed during the May 2014 pilot of the app, and (4) is a proposed deadhead or express alternative that turns down Dunster Street, skipping several stops around Harvard Yard. Versions (2) and (4) are similar in that they skip the same set of stops, but (2) takes a much longer distance to achieve the same result, and passes through six traffic signals. Therefore, we discard (2) as an inferior routing.

4.6.2 Evaluation of possible strategies

To evaluate possible strategies, we must examine the impacts on different groups of passengers. Each of these strategies saves time for a vehicle that is running behind, at the cost of additional waiting time for passengers at the stops that are skipped, as well as irritation

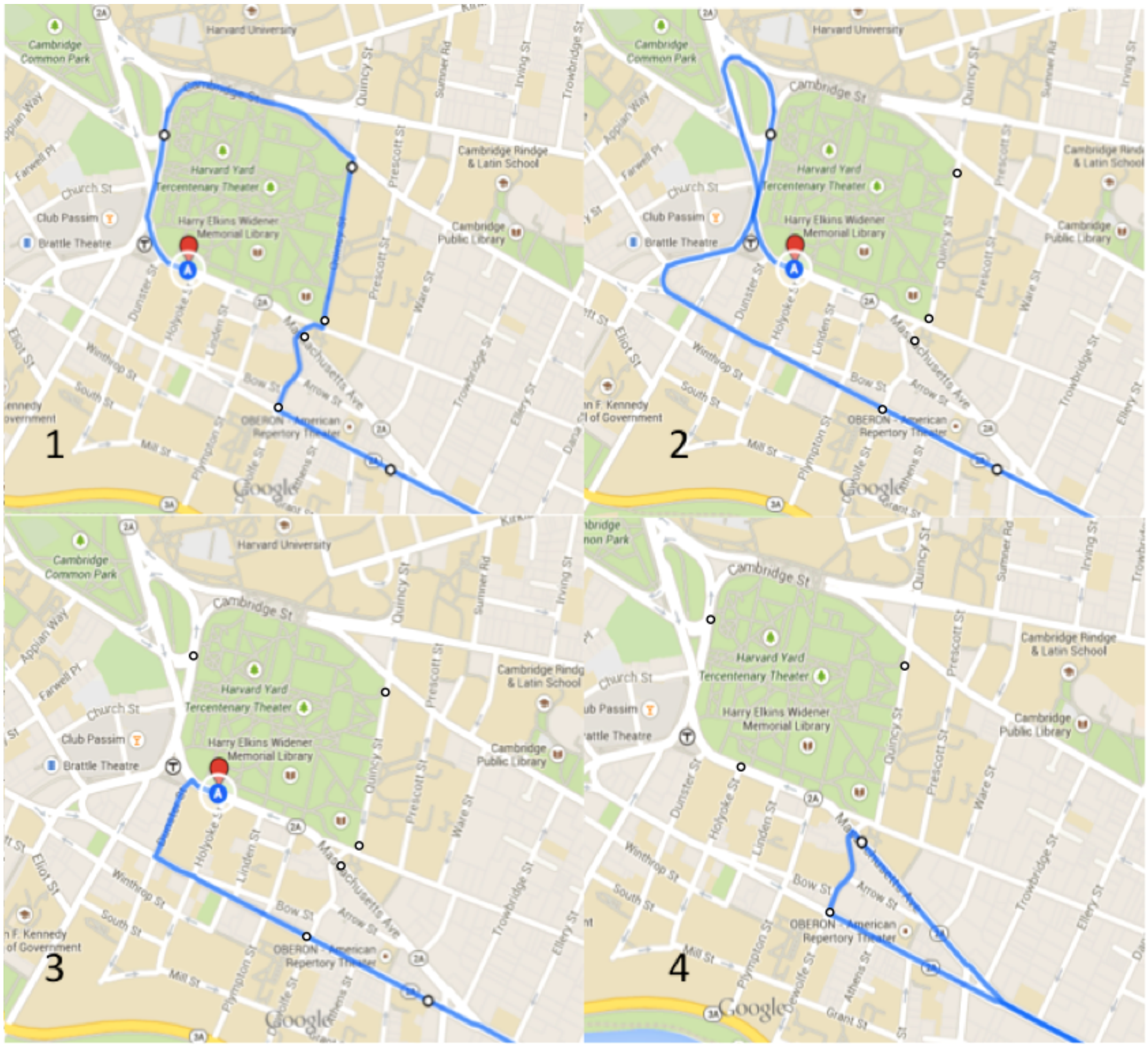

Figure 4-15: Four possible turnaround options at Harvard

from passengers who must alight early, or who see a bus go by without stopping.

Table 4.8 summarizes the impact on passengers by considering a case in which a bus has been delayed so that it is arriving with a leading headway double the regular scheduled headway. This is a typical example of a scenario where one of these strategies would be used. Using APC data on passenger boardings and alightings in the PM Peak period, we estimate the number of passengers who would be affected by each strategy in this scenario. Dwell-time impacts are estimated using a value of 3.5 seconds per boarding and 2 seconds per alighting (based on the Transit Capacity and Quality of Service Manual), and assuming that alightings are evenly split between the front and rear doors.

We distinguish between passengers who are denied boarding and those who are "bypassed completely," meaning that the bus did not pass by their stop at all. Passengers who are bypassed in this way will experience the longer waiting time but not the irritation caused by seeing a bus go by without stopping.

Table 4.8: Expected impact on customers of each strategy after a 2x-headway wait

Expected impact per trip	Express via Dunster St.	Deadhead via Dunster St.	Short-turn via Bow St.
Time savings	5:20	5:42	7:31
No. of passengers denied boarding	0	6.3	0
No. of passengers bypassed completely	13.6	13.6	20.0
No. of passengers forced to alight early	0	0	16.8

This analysis is based on a single scenario, but it makes clear the tradeoffs between deadheading, expressing, and short-turning:

- The short-turn saves about one additional minute of running time, at the cost of forcing all passengers destined for the Harvard stop to alight early.
- The deadhead option removes the need for passengers to alight early, but the passengers waiting at the terminal who would otherwise have been bypassed by the short-turn are denied boarding instead.
- The express option is similar to the deadhead, but passengers at the terminal are allowed to board at a small cost in additional dwell time.

The improvement in dwell time of the deadhead option over the express option is limited, due to the relatively small number of boardings that typically happen at the Harvard stop. Only if a large number of passengers were waiting at the stop, and the following bus were close behind, would it make sense to use the deadhead strategy. The short-turn strategy has the dwell-time advantage of not picking up Harvard passengers, as well as an estimated 1 minute and 49 seconds of additional running-time savings due to the shorter route. This time savings benefits all passengers waiting downstream, but may be reduced by the time needed to explain to passengers that they must alight early. It also must be weighed against the inconvenience and irritation caused to affected passengers, factors which will be discussed further below.

One potential issue with the Dunster Street route is its width. Although currently coaches do use Dunster Street, there may be scenarios in which a double-parked delivery vehicle could prevent a bus from passing down the street. The feasibility would have to be investigated through observations of Dunster Street and test runs, and potentially parking spaces on the street might have to be moved.

4.6.3 Low-visibility strategies

In general, passenger irritation is an important consideration in the costs of all of these strategies. Passengers expect to be able to board the first bus on their route that appears, and for it to bring them to their destination without stopping except to pick up and drop off passengers. Violating these expectations causes frustration and confusion among passengers beyond that caused by regular delays. Carrel et al (2013) studied factors that lead customers to reduce their transit ridership. They found that in-vehicle delays due to traffic, medical emergencies, or other factors immediately visible to the customer had no significant impact on a person's likelihood of reducing their transit use, but delays due to "problems downstream," including holding of vehicles, were associated with a strong and significant increase in the likelihood of reducing transit use.

Because of this poor perception of delays caused by visible control actions, agencies typically avoid measures such as expressing and midpoint holding except in extreme cases. Pangilinan et al (2008) conducted an experiment with the CTA in which expressing and deadheading were completely left out to avoid irritating passengers. Similarly, experiments by Bartholdi and Eisenstein (2012) and Strathman et al (2001) used holding at terminals only.

However, a short-cut such as the Dunster Street route discussed here enables a bus to skip stops without directly driving past them. If the express version of this route is used (allowing boardings at Harvard), the only passengers negatively impacted will be unaware that they have been skipped.

Table 4.9: Time from Harvard to 1st Timepoint (Mt Auburn St @ Putnam Ave) (minutes)

TimePeriod	Mean Time to 1st Timepoint	Std. Dev. Time to 1st Timepoint
AM Peak	7:03	1:45
Midday Base	6:53	2:07
Midday School	7:03	1:21
PM Peak	7:20	2:00
Evening	6:17	1:13
Other (Late/Early AM)	7:20	2:00

4.7 Recommendations

Based on the experiment results, we have identified above various problems that caused ineffective implementations of the control strategy, particularly at Dudley. We have also identified potential solutions to these problems. In this section, we make recommendations to the MBTA based on our observations from the experiment.

4.7.1 Departure discipline

Operator departure discipline at Dudley was identified as a significant problem. To address this problem, we recommend several measures based on those used by other agencies:

- A report of operator on-time performance for terminal departures similar to that shown in Section 4.5.1 should be generated on a regular basis, and used for conversations between garage managers and bus operators.

- The TransitMaster feature used by Metro Transit to alert dispatchers to early departures by bus operators should be enabled in the MBTA control center, and dispatchers should contact operators who appear to have departed early.

- The LED sign at Berth 20 at Dudley Station should be used to display the next scheduled departure time for Route 1.

4.7.2 Operations planning

Our examination of half-cycle times showed that insufficient time is provided for buses to complete trips on time with 95% confidence in the Midday School, Midday Base, and PM Peak periods. We recommend that the MBTA re-examine the scheduled cycle times on Route 1 to ensure that sufficient recovery time is provided in each direction. At certain times, this may be a matter of re-allocating cycle time from one direction to the other, while at other times the problem is in both directions, and only additional resources or lengthening scheduled headways could solve the problem.

4.7.3 Back-door boarding

Long boarding times at Dudley were identified as a problem. A CharlieCard validator should be provided to inspectors at Dudley, and should be used at the discretion of the inspector to allow back-door boardings on the Silver Line. Use of this device should be formalized and extended to Route 1 (and other routes easily accessible from the Silver Line berth); cutting down a long dwell time by allowing back-door boarding could have a large impact on an individual route's performance, while only requiring a few minutes of an inspector's time.

4.7.4 Layovers

Crowding in the bus lane and layover area at Dudley was identified as a problem. Bus operators should be instructed to park in the departure berth itself if they are next in line to depart, as in the procedure described in Section 4.5.3. Part of the layover area should be reserved for use as a passing lane so that Route 1 buses laying over at the berth will not prevent other buses from pulling through, as shown in Figure 4-14.

4.7.5 Guidelines for special control strategies

Special control strategies such as deadheading, expressing or short-turning are currently used on an ad-hoc basis by inspectors. The lack of a standardized strategy was a problem during the experiment, with one inspector implementing an ineffective strategy at Harvard. As a first step, standardized strategies should be identified whenever possible and taught to inspectors as part of their training. Special attention should be paid to "low-visibility" strategies such as the express route via Dunster Street. In the future, more research should be conducted on these strategies, including simulation testing, to lead up to the eventual incorporation of these strategies into an automated decision tool.

Chapter 5

Simulation

Experimental testing provides us with valuable information about the performance of control strategies under real-world conditions, but it is limited by cost and the willingness of transit agencies to allow their standard procedures to be changed for research purposes. Simulation modeling allows us to expand our analysis of alternative strategies by testing a wider range of control strategies under different conditions. In this chapter, we will describe a simulation model to evaluate a variety of possible control strategies, as well as different contexts in which they could be used.

The simulation model is based on work by Gabriel Sanchez-Martinez. The simulation framework was developed in his Master's thesis (Sanchez-Martinez, 2012), and implemented in various forms in his Ph.D. thesis (Sanchez-Martinez, 2014). The framework describes a method of simulating a high-frequency bus service using various automatically-collected data as inputs.

Sanchez-Martinez used two different types of model: A model of an idealized route, in which vehicles enter service at the beginning of the simulation and leave service at the end, and order of vehicles is preserved throughout the simulation, and a realistic route, with vehicle schedules drawn from a London Buses route. The idealized model was used to compare static and dynamic optimization strategies with headway-based holding strategies such as those used in this research. The realistic model was used to test the dynamic optimization strategy in various contexts. In this research, we adapt the realistic route model to use MBTA data sources as inputs, and to simulate holding and other strategies in the context of the MBTA's vehicle schedules. We also add to the model a more detailed treatment of operator departure behavior at terminals, based on the work of Milkovits (2008).

5.1 General framework

5.1.1 Route

A *route* consists of (typically) multiple *variations*. Each variation consists of a sequence of locations, in the order that they are visited by vehicles serving the route. A simple uni-directional loop route might have only one variation, while a simple bi-directional route will have one variation for each direction. Multiple variations exist when a route has branches or different locations where vehicles might reverse direction or enter (leave) service.

5.1.2 Locations

Locations in this model represent terminals, stops, and route segments. Each location is assigned a *location controller*, which governs the behavior of vehicles at that location. When a vehicle arrives at a location, the location controller determines how long the vehicle will remain at that location, what location the vehicle will visit next, and how many passengers will board or alight at that location.

- *Stops* - A stop is a location at which passengers may board or alight from a vehicle. Each stop controller draws from a Poisson distribution, using the arrival time at the stop of the previous vehicle, to determine the number of passengers waiting to board the vehicle. The number of passengers alighting is determined based on an "alighting fraction", that is, the fraction of passengers that typically alight at the stop, by variation and time period. It determines the next location for the vehicle based on the route variation being served. The time that a vehicle departs is determined using control logic. For example, a holding strategy may be applied or a vehicle may simply be dispatched immediately.

- *Terminals* - A terminal is a location where vehicles may be brought into service or taken out of service. A terminal is considered distinct from a stop, even in cases where a stop and a terminal may occupy the same physical space. A terminal controller determines, based on the schedule, whether a vehicle arriving at the terminal should go out of service. If it continues in service, the controller assigns its next scheduled trip, and determines the dispatch time and next location based on the variation being served.

- *Links* - A link represents a segment of the route between two stops. Each link controller determines the running time of the vehicle on that link. In our simulation, running times are drawn from a bivariate distribution of observed running times, as described below.

5.1.3 Vehicles

A *vehicle* is an object with a specific identity in the simulation. Each vehicle is assigned to a *block* from the schedule, giving it a series of scheduled trips to run on the different variations of the route. Vehicles are the only type of agent in this simulation, meaning that passenger boardings and alightings, for example, are handled by the vehicle object. This simplifies the tracking of costs to on-board passengers.

5.1.4 Events

The model used here is an event-driven simulation, meaning that events such as vehicle arrivals at stops and passenger boardings and alightings are processed in chronological order. A *heap* data structure is used to store events as they are generated and process them chronologically. A *replication* is a single run of the simulation, representing one day of service on the modeled route.

5.1.5 Terminal departure behavior

The initial departure time of a vehicle on a trip may be affected by a variety of factors. There are four distinct processes which may affect terminal departures, each of which is subject to operator behavior and external sources of randomness.

- *Recovery time* is the period between the arrival of a vehicle (and operator) at a terminal, and its subsequent departure on its next scheduled trip. This time is often used by operators for breaks, leading to variability in the amount of time taken.

- *Garage pull-outs* happen before the first trip of the day for a particular vehicle. Variability in these first departures may be caused by delays leaving the garage, or by traffic between the garage and the start of revenue service.

- *Operator reliefs* generally take place at terminals, where one operator ends their shift and their vehicle is taken over by a new operator. The major source of delay from operator reliefs is late arrival by the relieving operator.

- *Interlining* is the practice of scheduling a single operator to run on multiple routes during a single shift. Since our model only simulates a single route, this means that some vehicles enter the simulation having previously arrived at a terminal from a different route.

The topic of terminal departure behavior was explored extensively by Milkovits (2008). We follow his method of simulating recovery time as the maximum of two values: Minimum required recovery time, and available recovery time. Minimum required recovery time represents the time needed for a vehicle operator to take a personal break. This is stochastic, and modeled using a normal distribution. Available recovery time is the amount of time available before the vehicle is supposed to depart on its next trip, whether based on schedule or following a control strategy. This value has a random component representing the operator's deviation from the assigned departure time. It is modeled as a two-stage distribution: First, a uniform distribution is used to determine whether the departure will be early or late, and then an exponential distribution determines the magnitude of the schedule deviation. The estimation of parameters for these distributions is discussed below.

In addition to the operator behavior modeled by Milkovits, we must also consider interlining and garage pull-outs, two situations in which a bus enters the simulated system with a degree of randomness. We handle this by drawing from the distribution of observed arrival times at the first terminal departure of the day, for each vehicle. Once the vehicle has arrived at the terminal, we use the minimum recovery time and schedule deviation as described above to determine when it will depart.

5.1.6 Boardings and alightings

Passenger boardings and alightings are modeled in this simulation to allow for measurement of the impact on passengers of the different strategies to be tested. Since we are not explicitly modeling dwell time, the number of passengers waiting at a stop will have no impact on the running time of a vehicle. Passenger arrivals at a stop are generated using a Poisson process, while alightings at each stop are determined using the *alighting fraction*, (as defined above).

5.2 Adaptation to MBTA context

5.2.1 Routes

For this research we will model two MBTA bus routes. Routes 1 and 28 are both part of the Key Bus Routes program, an MBTA initiative to emphasize service quality on high-density corridors with frequencies and spans of service similar to the MBTA rapid transit services. Route 1, discussed in Chapter 4, is a cross-town route connecting various residential and commercial neighborhoods along its route from Harvard Square to Dudley Square. Passengers Route 28, by contrast, is a radial route connecting Mattapan, a primarily residential section of Boston distant from downtown, with Dudley and Ruggles Stations, two major transit hubs closer to the employment centers of the city. Load profiles on the two routes are quite different, as seen in Figures 5-1 and 5-2. Route 28 has loads that gradually increase from Mattapan in to Dudley, and strongly directional travel patterns (mainly inbound loads in the AM Peak, and outbound in the PM Peak), while Route 1 has its maximum load point near the middle of the route, and roughly even loads between the two directions of travel.

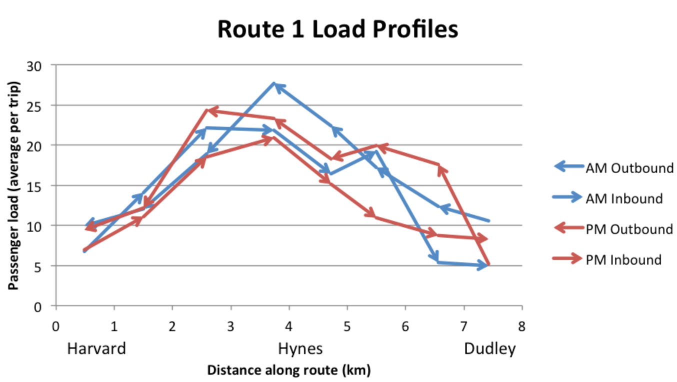

Figure 5-1: Load profiles on Route 1 (Fall 2014)

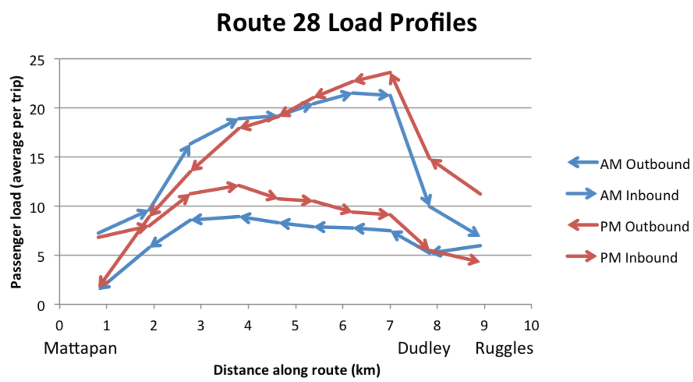

Figure 5-2: Load profiles on Route 28 (Fall 2014)

5.2.2 Segment running times

Running times are drawn from an observed distribution of running times taken from the MBTA's Automatic Vehicle Location (AVL) system. This system uses a "geo-fencing" technique to measure arrival and departure times at major stops, or *timepoints*, along each bus route. We limit our simulation to timepoints and the route segments between them, rather than including all stops on the route, in the interests of simplicity and because the geo-fencing technique creates the most accurate data available from the MBTA's AVL system on arrival and departure times. The data are taken from weekdays during the period from September 15, 2014 to October 31, 2014, excluding Fridays, which have a different schedule from the other weekdays, and the Columbus Day holiday. The sample thus contains 27 days of observations. Running times on each link are drawn from a bivariate distribution similar to that described by Sanchez-Martinez (2012). First, observed running times on a link are grouped into 30-minute time periods, e.g. 5:30 AM-6:00 AM. Next, within each time period, each running time observation is paired with the observed running time of the same vehicle on the previous segment. On the initial segment of a simulated trip, a running time is drawn randomly from the observations in the 30-minute time period. On subsequent segments, the window is further narrowed based on the running time from the previous segment. For example, if a simulated bus is departing Harvard at 9:40:00 AM, it will draw a running time on the initial link from the set of observations between 9:30 AM and 10:00 AM. Assuming it draws a value of 384 seconds, it will arrive at the Putnam timepoint at

9:46:24 AM. On the following link, the running times will first be narrowed down by time period (in this case the same time period as before), and then by previous running time, using a window width equal to 20% of the previous running time, in this case 76.8 seconds. So only observations taken between 9:30 AM and 10:00 AM, for which the previous running time was between 345.6 and 422.4, will be available to draw from. The size of the windows from which running times were drawn was examined by running the simulation 100 times, and observing each window that was sampled. For segments that were the first in their trip, the window sizes ranged from 5 to 103, with a median of 72. For subsequent segments, the windows narrowed down by previous running time were found to range from 1 to 112, with a median of 18 observations.

5.2.3 Terminal departure behavior

Operator behavior departing a terminal is a key component of transit operations. As discussed in Chapter 4, operator adherence to schedule or to supervisor instructions can make a significant difference in the effectiveness of a particular control strategy. Therefore, it is necessary to add a component to our simulation to model the behavior of operators at terminals. We build upon the work of Milkovits (2008), who examined the behavior of operators on a high-frequency bus route in Chicago.

Minimum recovery time

To create a distribution for minimum recovery time, we examine the AVL dataset for cases in which no recovery time was available, that is, the arrival time at the first timepoint of a trip was later than the scheduled departure time. In these cases, we assume that bus operators will take the minimum possible recovery time, given their constraints including passenger unloading and loading, as well as restroom or other breaks. We exclude cases where the recovery time was in excess of 20 minutes; these are likely outliers caused by mechanical failures or other issues unrelated to operator behavior.

Table 5.1 shows summary statistics of the recovery times. As was noted in Chapter 5, less slack time is scheduled for Route 1 buses at Harvard than at Dudley, but recovery times at Harvard are shorter and less variable. On Route 28, the number of trips with no available recovery time is more evenly balanced between the two terminals, but recovery times are both longer and more variable than on Route 1.

Table 5.1: Minimum required recovery time distribution

Route	Terminal	Observations	Mean	Std Dev	Min	Max
1	Dudley	553	3.1	2.4	0.32	18.0
1	Harvard	755	2.3	1.6	0.28	18.0
28	Ruggles	603	3.5	2.2	0.85	19.3
28	Mattapan	563	4.5	2.8	0.13	9.6

Schedule deviation

In order to measure schedule deviations, we will need to examine a sample of departures for which sufficient recovery time was available for an operator to depart on-time. To set a cutoff value that defines "sufficient recovery time," we examine the cumulative distribution function of recovery times taken when a vehicle arrived after its scheduled departure time in Figure 5-3.

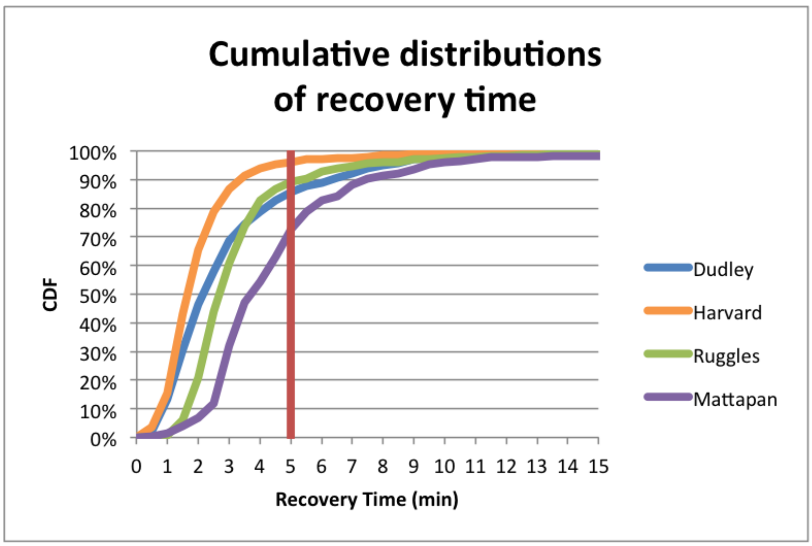

Figure 5-3: Cumulative distribution of actual recovery times for buses arriving with no available recovery time

Notably, the minimum recovery time that was required at Mattapan was significantly higher than at the other terminals. In order to determine the cause of this, field observations were conducted of bus departures on Routes 28, 31, and 245 (which have berths in the same section of the station) at Mattapan. Comparison of the observed departure times with AVL data suggested that part of the variability in departure times is due to a traffic signal that these buses pass through to exit the station onto Blue Hill Avenue. This signal has a two-minute cycle with 106.5 seconds of red time and 13.5 seconds of green time. Because they are still within the geo-fence area while stopped at the signal, the buses' departures are not recorded by the AVL system until they pass through the signal. Similar issues may exist at Ruggles station, where Furth et al (2010) found that signal delay was a major factor impacting bus service around the terminal. The remainder of the high variability at Mattapan seems to be due to bus operators' tendency to take long breaks due to the ample break facilities available, nearby amenities such as coffee shops, and lack of supervision.

Milkovits used a cutoff value of 4 minutes to include at least 80% of observed minimum recovery times at each terminal. In our case, the 80th percentile of minimum recovery times at Mattapan would be 6 minutes, but we subtract one minute to account for the added variability due to the traffic signal, as this is present in all cases regardless of available recovery time. Using our cutoff value of 5 minutes, we fit an exponential distribution to the early and late departures from each terminal, shown in Figures 5-4 and 5-5. In general, the exponential distribution fits the observed deviations well. The parameters for the distributions used are given in Table 5.2.

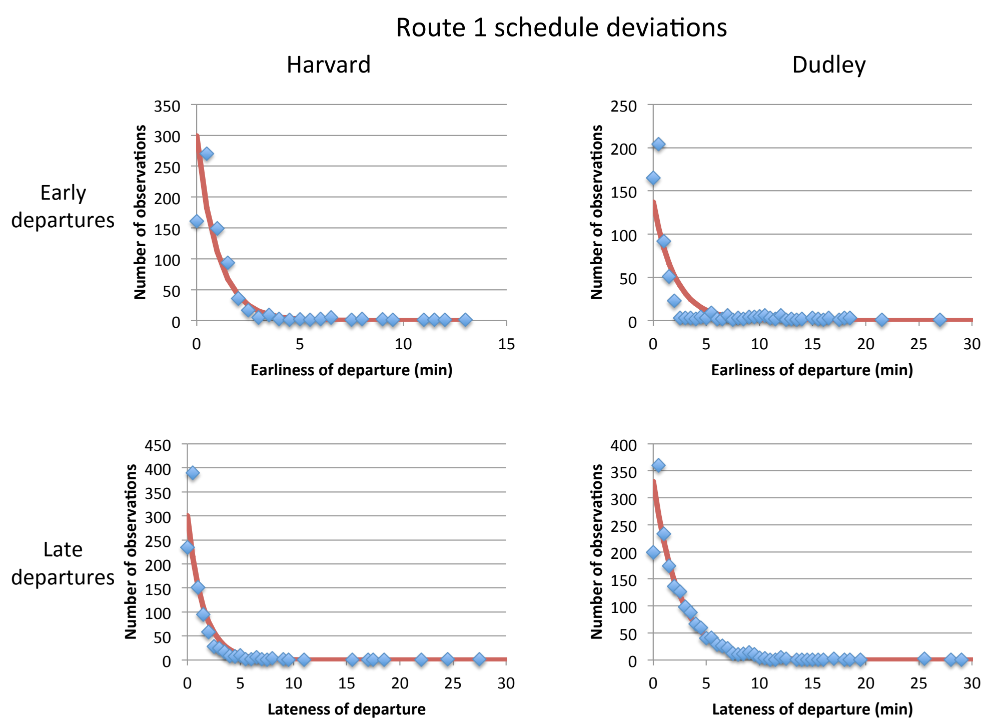

Figure 5-4: Route 1 - Schedule deviation among vehicles with >6 minutes of recovery time available

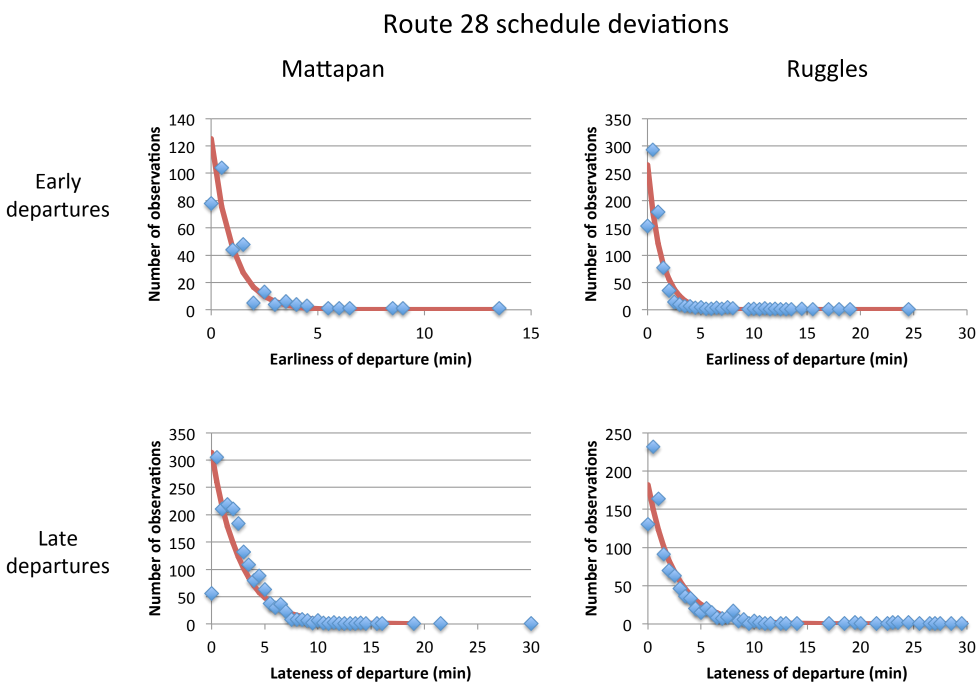

Figure 5-5: Route 28 - Schedule deviation among vehicles with >6 minutes of recovery time available

Table 5.2: Schedule deviation - distribution parameters

			Early departures		Late departures	
Route	Terminal	Observations	Mean earliness	% Early departures	Mean lateness	% Late departures
1	Dudley	2643	119	26%	148	74%
1	Harvard	2081	61	41%	91	59%
28	Ruggles	2091	74	44%	148	56%
28	Mattapan	2346	58	15%	166	85%

5.2.4 Even-headway strategy

The minimum-layover and schedule-deviation factors that were modeled in Section 5.1.5 create deviations from the scheduled departure times, an important component of the simulation. With an even headway strategy, we expect to observe similar deviations from the departure times after holding. Indeed, we observed in the experiment results in Chapter 4 that departure times frequently deviated from the instructions given to operators and supervisors. In this section we will refer to the deviation factor more generally as "departure deviation," to include both the schedule-following and even-headway strategies.

For the even-headway strategy at terminals, we will draw values from the same distributions that were used for calculating minimum layover time and departure deviation in the schedule-following case. Minimum layover time functions in the same way: Upon a vehicle's arrival at a terminal, a value is drawn for the minimum required layover at that terminal, and the vehicle is not permitted to leave before the layover is up.

However, departure deviation is somewhat more complicated. The key distinction between the even-headway strategy and the schedule-following strategy is that in the even-headway strategy, departure times are updated as new information becomes available. Specifically, in our simulation the departure time is updated every minute. The random values for departure deviation cannot be re-drawn each time the holding period is re-calculated, as this would bias the distribution of deviations against extreme values.

A naïve solution to this problem would be to draw the value for departure deviation only once for each terminal departure, and to re-use that same value even when the holding period is re-calculated. This creates a problem, however, when a particularly long value is drawn. For example, if a deviation of 10 minutes of lateness is drawn during a period when scheduled headways are less than 10 minutes, the vehicle will sit at the terminal for the full 10 minutes, during which time one or more other vehicles will depart. It will then re-calculate a new, later departure time, and add 10 minutes to that. Vehicles that draw long departure times will thus get "stuck" at the terminal.

To solve these problems, while retaining the ability of the even-headway strategy to update its departure times periodically, we use the following procedure upon the arrival of a vehicle at a terminal: 1. Draw a value for minimum layover time (as in Section 5.2.3) 2. Hold until the minimum layover time has passed 3. Draw a value for departure deviation (as in Section 5.2.3) a. If early departure: Set the earliness as a fixed value for this vehicle. All recalculated departure times will be adjusted by this earliness value. b. If late departure: Add the lateness value to the initial calculated departure time. This departure time becomes a minimum departure time; the vehicle may not depart before this time. If the departure time is re-calculated to fall after this minimum, the vehicle will depart at the exact re-calculated time.

This procedure can be explained conceptually as follows: Early departures can be thought of as being driven by an operator's desire to stretch the rules by departing earlier than instructed, and thus may move earlier or later as the departure is recalculated. Late departures, on the other hand, are typically caused by a vehicle operator being away from the vehicle at the instructed departure time. The value that is drawn for lateness can be thought of as the amount of time before the operator returns to the vehicle. Upon the operator's return, he is able to depart at the instructed time, and thus we do not add more lateness if the departure time is recalculated.

5.2.5 Boardings and alightings

Passenger boarding and alighting information is derived from two sources: Automated Fare Collection (AFC) and Automated Passenger Counter (APC) data.

AFC data is used to estimate passenger arrival rates at stops with the following procedure: The day is divided into 30-minute time periods. For each stop, the number of boardings on each bus is determined, along with the leading headway. Boardings on the first bus of each period are split between that period and the preceding period, proportionally to the part of the headway that was in each period. Passengers boarding the first trip of the day at a stop are assumed to have arrived randomly during the five minutes prior to the scheduled departure of the bus from that stop.

APC data is used to estimate the alighting fraction. The APC system determines a passenger load and number of alightings for each stop, and the alighting fraction is averaged over all alightings at a stop during each 30-minute time period.

Stop-level boardings are aggregated to the next timepoint along the route. This is done so that the sum of boardings will equal the estimated load at each timepoint.

5.3 Validation

In order to verify that the model sufficiently represents the transit system being studied, we must go through a process of *validation*. To validate our simulation, we attempt to simulate the conditions of typical MBTA service, and compare our results to observed data.

In this section we compare the performance of buses on a simulated version of MBTA Route 1 with the actual performance on that route from September 15, 2014 to October 31, 2014. For evaluating performance statistics, the MBTA divides the service day into 7 time periods, shown in Table 5.3. The time periods of interest to our simulation are the AM Peak, Midday Base, Midday School, and PM Peak periods, as these are the busiest periods of the day and have high frequencies of service, making them suitable time periods for the implementation of headway-based control strategies. We follow a similar method to Sanchez-Martinez (2012), and calculate each statistic over groups by time period and location, plotting them against a 45-degree line to visualize the similarity or difference, and calculate the root mean squared error (RMSE) of the simulation results with respect to the observed data.

Table 5.3: Time periods for analysis

Time Period	Start	End
Early AM	06:00	07:00
AM Peak	07:00	09:00
Midday Base	09:00	13:30
Midday School	13:30	16:00
PM Peak	16:00	18:30
Evening	18:30	22:00
Night/Sunrise	22:00	06:00

5.3.1 End-to-end running time

Figure 5-6 shows a comparison of means and standard deviations of end-to-end running times in the simulated dataset, as compared with the real-world observations. One dot represents a single direction of travel during one of the time periods. Our simulated mean values of end-to-end running times match up very well with the observed values. The RMSEs are 3.1 minutes for Route 1 and 2.1 minutes for Route 28, which is fairly low given that the mean values were between 35 and 55 minutes. The simulated variability in end-to-end running times was somewhat less accurately modeled, as can be seen on the lower two charts. Variability of running times was underestimated by the simulator on average, with an RMSE of 2.0 minutes for Route 1, and 1.6 minutes for Route 28. Our simulation takes into account correlations between adjacent segments, as described in Section **??**. However, it is likely that more complex correlations exist that drive running-time variability higher than our simulator suggests.

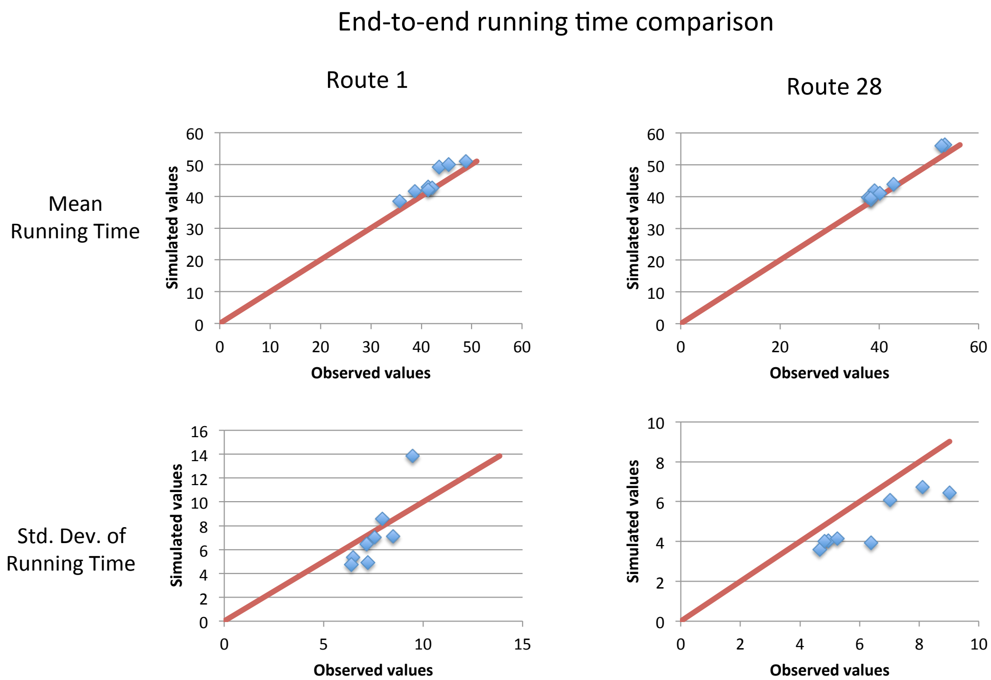

Figure 5-6: Validation of end-to-end running time means and standard deviations

5.3.2 Headways

Figure 5-8 shows the comparison of headway statistics from the simulator results against the observed data. Each point represents a timepoint in a particular direction during a time period. There is a clear distinction between the results for Route 1 and Route 28 in this measure. The mean headway values from the Route 1 simulation match the observed values very closely, with an RMSE of 0.5. The standard deviations are somewhat less closely matched, but show no particular bias.

The mean and standard deviation values for Route 28, on the other hand, are consistently underestimated by the simulation. The lower mean-headway values indicate that fewer trips are being made on Route 28 in real life than in the simulator. To test this, we compared the scheduled number of trips, as well as the scheduled number of runs, with the actual trips and runs served on each day in the sample (see Figure 5.4). On Route 28, a median of 7.3% of scheduled trips were dropped each day, and a median of 1 scheduled run was dropped completely. These dropped trips could be due to operator absenteeism, vehicle maintenance issues, or vehicles being moved to serve other routes.

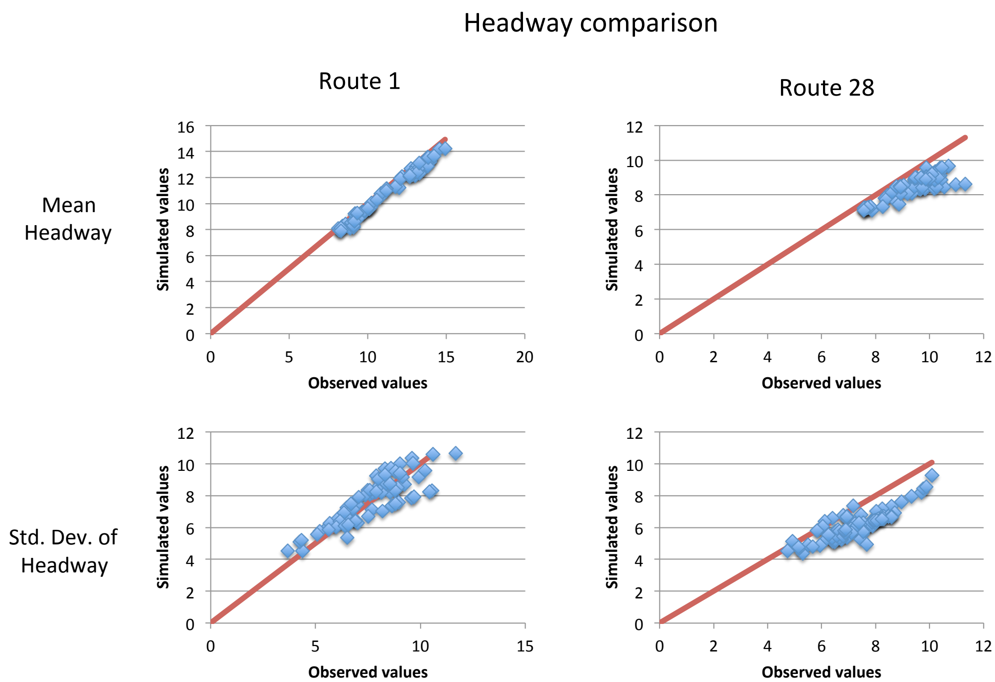

Figure 5-7: Validation of headway means and standard deviations

Table 5.4: Dropped trips and runs

	Dropped trips		Dropped runs	
	Route 1	Route 28	Route 1	Route 28
Minimum	0 (0%)	1 (0.4%)	0 (0%)	0 (0%)
Median	5 (2.2%)	17 (7.3%)	0 (0%)	1 (2.6%)
Maximum	13 (5.8%)	31 (13.3%)	2 (4%)	3 (7.7%)
Total scheduled	223	233	50	39

5.3.3 Even-headway strategy

The simulation of the even-headway strategy, as described in Section **??**, is based on the strategy used in the experiment described in Chapter 4, but uses the same distributions of departure-time deviations, rather than using the distributions observed in the experiment. The purpose of this is to provide an "apples-to-apples" comparison of the two strategies, with similar levels of deviation from instructed departure times, and to allow for the effect of reducing departure-time deviations to be studied separately from the effect of the strategy. In addition, the experiment provided a relatively small sample of data, with only four weekdays with one four-hour period each weekday. Due to the fact that the sample size is small and the model is based on driver behavior observed during the schedule-following strategy, we do not necessarily expect a close correspondence between the simulation results and the observed data.

Table 5.5 shows a comparison of running-time statistics from the experiment and from the simulation of the even-headway strategy on Route 1. (The "Midday" time period referenced here is that portion of the Midday School period that was part of the experiment, from 2:30 PM to 4:00 PM). Running times were notably both lower and less variable during most of the experiment time periods than in the simulation. This may be partly due to the implementation of the strategy leading to faster running times through dwell-time factors not modeled in the simulation, but the sample size is small enough that it may simply be due to random variation in travel times.

Figure 5-8 compares the simulated and observed values of mean headway and standard deviation of headway, aggregated at the timepoint and time-period level as in Section **??**. Observed mean headways were slightly longer than simulated, implying that fewer trips were completed during the experiment than in the simulation, and the Root Mean Squared Error of mean-headway values was 0.9 minutes.

The standard deviations of headways have a worse fit, with a RMSE of 1.5 minutes.

Table 5.5: Comparison of running-time statistics for even-headway strategy

Direction	Time Period	Mean Running Time		Std Dev of Running Time	
		Observed	Simulated	Observed	Simulated
From Dudley	Midday	39.9	42.1	3.0	5.7
	PM Peak	43.7	48.8	7.3	13.0
From Harvard	Midday	52.0	52.6	8.9	6.7
	PM Peak	48.3	50.1	6.3	8.8

In general, the variability of headways on trips from Harvard was better in the experiment than in the simulation, while the variability of trips from Dudley was at times worse, and at times better. This agrees with the observations from Chapter 4 that trips from Harvard performed significantly better than trips from Dudley during the experiment. The RMSE for trips from Harvard was 1.7 minutes, while for trips from Dudley it was 1.2 minutes.

The fact that trips from Harvard had better performance in the real-life experiment than in the simulation suggests that, as expected, the simulation is pessimistic when calculating departure deviations for the even-headway strategy. The relationship between departure deviation and the even-headway strategy will be explored further in the next chapter.

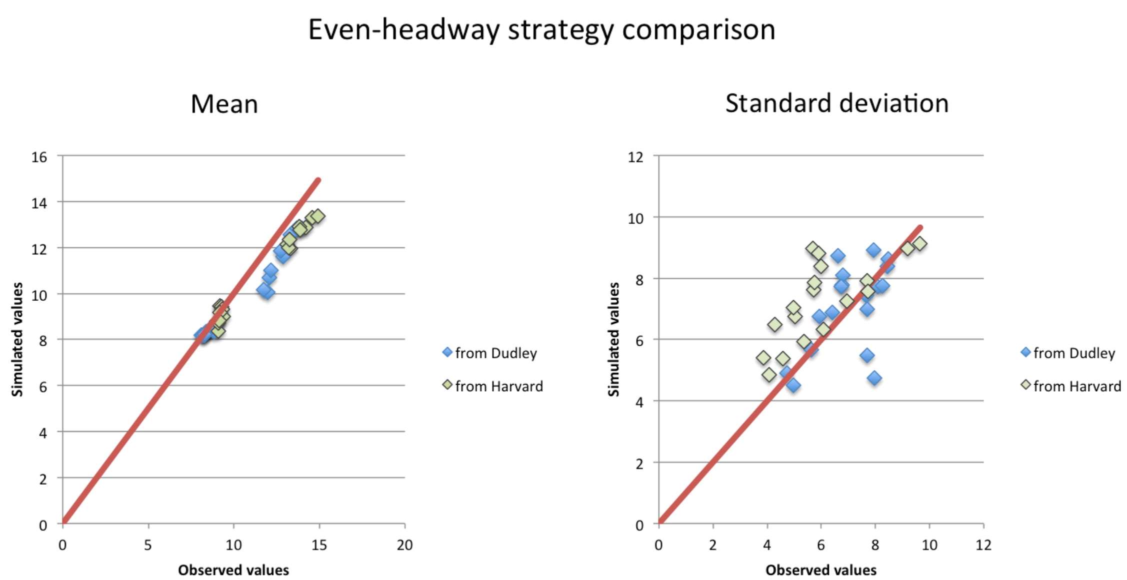

Figure 5-8: Comparison of headway statistics for even-headway strategy

5.3.4 Summary of validation

The root mean squared error statistics are summarized in Table 5.6. In general, by adding the element of operator behavior at terminals, we have improved upon the Sanchez-Martinez model in better capturing the mean and standard deviation of headways. The downward bias of the model in predicting values for Route 28 is concerning, but the RMSE values are nevertheless low. We were unable to achieve as close a fit to the distribution as Milkovits did, most likely because the Milkovits model explicitly incorporated dwell times and passenger behavior.

Table 5.6: Validation summary - root mean squared error statistics

	Running times		Headways	
	Route 1	Route 28	Route 1	Route 28
RMSE of means (minutes)	3.1	2.1	0.5	1.0
RMSE of std. devs. (minutes)	2.0	1.6	0.8	1.3

Chapter 6

Simulated Experiments

In this chapter, the simulation model developed in Chapter 6 will be used to further explore control strategies based on real-time vehicle location data. The simulation experiments will explore the following scenarios:

- The route on which the strategy is implemented
- The magnitude of deviations from assigned departure times
- The number and location of control points
- The type of strategy employed

We will first determine metrics to assess the results of the experiments, including both metrics based on the passenger experience, reflecting the wait times, as well as a metric that reflects the cost to agencies of the additional vehicles required to provide a good service despite irregular headways.

6.1 Passenger experience metrics

6.1.1 Passenger wait time

Passenger wait time is the time between a passenger's arrival at a stop, and the time that passenger boards a vehicle. We will consider both total passenger wait time (TPWT) of all passengers during a particular time period, as well as average passenger wait time (APWT). We use the following notation, taken from Tribone (2013):

λ_p^o = passenger arrival rate at origin station o in period p

h_i^o = headway for trip i at station o

Assuming that passengers arrive at a uniform rate (a random arrival process), the average wait at station o for trip i will be $\frac{h}{2}$, and the total number of passengers served by this trip will be λh. Therefore, the total passenger wait time over all stations and trips within a time period is as follows:

$$TPWT = \frac{1}{2}\sum_i \sum_o (\lambda_p^o h_i^o * h_i^o) = \frac{1}{2}\sum_i \sum_o \lambda_p^o (h_i^o)^2 \tag{6.1}$$

The average passenger wait time, then, is simply the total divided by the number of passengers boarding during the time period:

$$APWT = \frac{1}{2}\frac{\sum_i \sum_o \lambda_p^o (h_i^o)^2}{\sum_i \sum_o \lambda_p^o h_i^o} \tag{6.2}$$

Under "ideal" conditions of perfectly even headways, the APWT will be half the scheduled headway. Under real-world conditions, the APWT will always be higher than this value. For example, Route 1 during the PM Peak has a constant scheduled headway of 8 minutes, which under perfect schedule adherence would lead to an APWT of 4 minutes. However, the observed APWT for this time period on Route 1 between Sep. 15, 2014 and Oct. 3, 2014 was 6 minutes, 19 seconds.

6.1.2 Effective headway

Effective headway, as defined by Tribone (2013), measures the average headway at a stop, weighted by the number of passengers experiencing each headway. Equation 6.3 defines the effective headway at a stop, which is derived based on the assumption of random arrivals by passengers at a rate of λ. The number of passengers experiencing headway h_i is λh_i, so the effective headway H_E is as follows:

$$H_E = \frac{\sum_i (\lambda h_i^2)}{\sum_i \lambda h_i} = \frac{\sum_i (h_i^2)}{\sum_i h_i} \tag{6.3}$$

We may extend the effective headway metric to cover all stops on the route by weighting each observation by the passenger arrival rate at each stop (using the notation of the previous section):

$$H_E = \frac{\sum_i \sum_o \lambda_p^o (h_i^o)^2}{\sum_i \sum_o \lambda_p^o h_i^o} \tag{6.4}$$

Because of the assumption of a random arrival process for passengers, the effective headway is simply twice the average passenger wait time. It is a useful metric as it acts as a bridge between the passenger-experience perspective, represented by average passenger wait time, and the service-planning perspective, represented by scheduled headway. This relationship will be used in Section 6.2 to create a new metric relating headway variability to the resources needed to operate a route.

6.2 Additional Vehicles Required metric

Metrics based on the passenger experience, as described above, are useful in communicating the cost to passengers of irregular service in a way that is easy for passengers to understand. However, passenger experience is only one consideration faced by decision-makers at transit agencies; cost considerations are equally important in the decision-making process. An agency operating a route with irregular headways should consider the implicit cost consisting of the additional resources needed to operate its scheduled level of service. Thus, it is useful to translate the passenger-focused metrics into a resource-based metric that relates the irregularity of headways to the number of vehicles required to operate a route.

In this section, we define a metric, which we call "additional vehicles required," to approximate the number of additional buses that would be needed to operate the route such that the actual effective headway would be equal to the original scheduled headway, given no change in operations management and control. Using this metric, managers can measure the impact of an intervention in operations management not just in minutes of passenger wait time saved, but also as an increase in the efficient use of resources.

6.2.1 Assumptions

The Additional Vehicles Required metric is a translation of existing metrics using equations that relate service characteristics such as scheduled headway, scheduled cycle time, and coefficient of variation of headways. The relationships between these values will be true only if certain assumptions are satisfied, and may be thought of as "approximately true" in cases where these assumptions are not fully satisfied. The assumptions are as follows:

1. The arrival rate of passengers is constant over the time period at the selected stop.
2. The capacity of vehicles is non-binding, meaning that no passenger is denied boarding.
3. The process of vehicle arrival and departure is independent of the passenger arrival and boarding process.
4. The mean observed headway for the time period is the scheduled headway for the time period.
5. The scheduled cycle time, scheduled headway, and number of vehicles in service are constant over the time period.

Assumptions 1-3 are similar to the assumptions used to set up the simulation. Assumption 4 will be approximately true, given a sufficiently large sample size. We achieve a large sample size by running 500 replications of the simulation for each scenario. Assumption 5 depends upon the existing schedule for the service. Because of the fluctuating levels of service provided throughout the day, as well as the constraints imposed by vehicle and crew scheduling needs, and the MBTA's use of a scheduling system which makes fine adjustments to cycle times throughout the day, Assumption 5 will not be perfectly satisfied.

In order to select appropriate time periods for analysis, we examine the scheduled headways, cycle times, and number of buses in service throughout the day. In Figures 6-1 and 6-2, we show the scheduled cycle times measured beginning at the Boston terminal (Dudley or Ruggles) and the scheduled headways in each direction. Figure 6-3 shows the number of buses in service throughout the day on each route. Cycle times, headways, and buses in service vary significantly throughout the day, and do not follow the same patterns across both routes. The AM Peak, from 7:00 AM to 9:00 AM, is highlighted as a time period during which both routes have fairly consistent cycle times and headways, although both oscillate between headways that differ by one minute. The number of buses in service is also consistent for each route, apart from a "run-up" period at the beginning of the time period during which new buses are coming into service. Route 28 also has a long period in the afternoon with fairly consistent scheduled headways and cycle times, and 13 buses in service. In the analysis of the simulation, we will use the AM Peak period for both routes, and the Midday School period on Route 28, for illustration of the "additional vehicles required" metric.

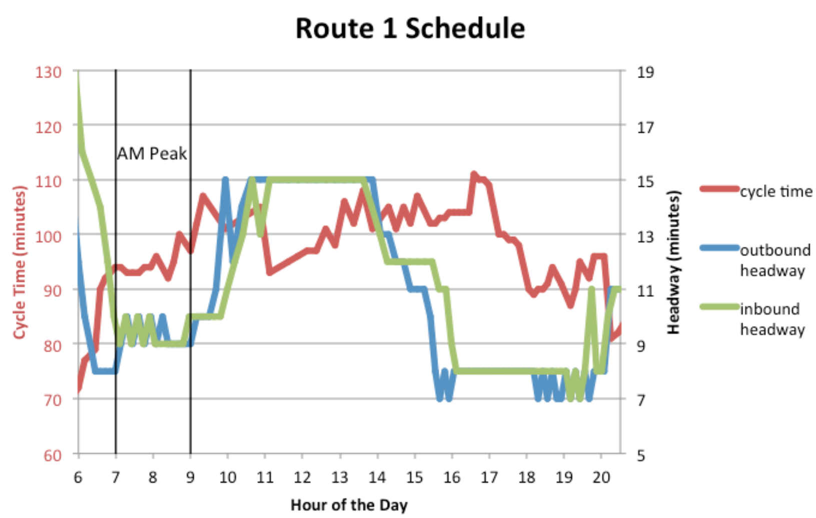

Figure 6-1: Scheduled cycle times and headways on Route 1

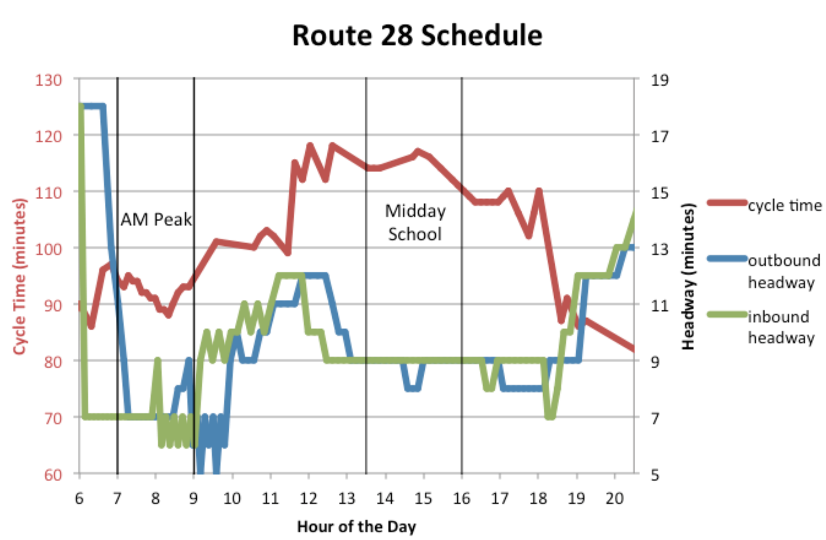

Figure 6-2: Scheduled cycle times and headways on Route 28

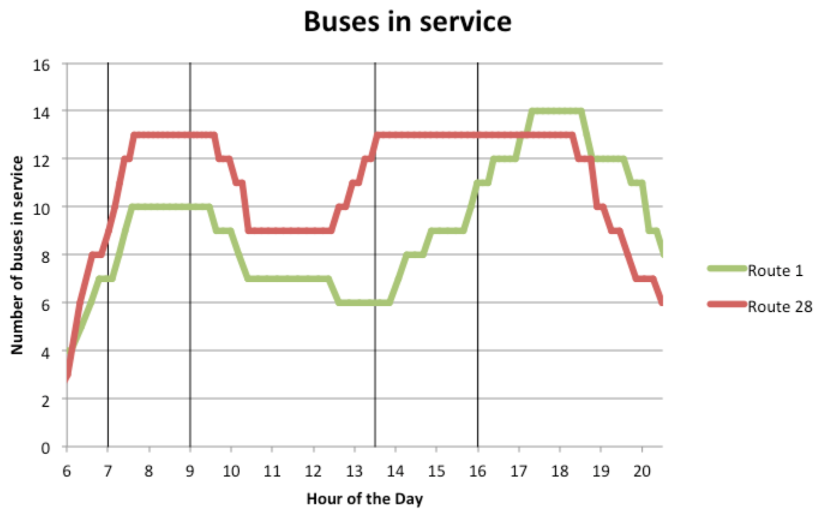

Figure 6-3: Number of buses in service per Fall 2014 schedule

6.2.2 Formulation of metric

We begin with an alternative formulation of the effective headway metric, shown in Equation 6.5 with $\hat{\mu}_h$ representing the sample mean headway and $\widehat{CV}_h^2$ the sample coefficient of variation. This is a version of the formula described by Welding (1957), and, as shown in Appendix A, is equivalent to Equation 6.3.

$$H_E = \hat{\mu}_h(1 + \widehat{CV}_h^2) \tag{6.5}$$

In Appendix B we show that this formula can be expanded to cover all stops along the route by weighing each stop by the passenger arrival rate at that stop using the following definitions of weighted mean and weighted coefficient of variation (for convenience, define $M = \sum_o \sum_i \lambda_p^o$):

$$\hat{\mu_h} = \frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o \tag{6.6}$$

$$\widehat{CV}_h = \frac{\sqrt{\frac{1}{M} \sum_i \sum_o \lambda_p^o (h_i^o)^2 - \left(\frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o\right)^2}}{\frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o} \tag{6.7}$$

In addition to this formula, we use the simple formula $c_S = nH_S$, which shows that the scheduled cycle time c_S on a route is equal to the scheduled headway H_S times the number of buses operating on the route n.

For this metric, we use the subscript 0 to represent the existing schedule, and the subscript 1 to represent the hypothetical case where additional buses are added to bring the effective headway H_{E_1} down to the level of the original scheduled headway H_{S_0}. Using Assumption 4, we substitute the scheduled headway H_{S_0} for the mean headway $\hat{\mu_h}$ in Equation 6.5:

$$H_{E_0} = H_{S_0}(1 + \widehat{CV}_0^2) \tag{6.8}$$

With our assumption that the scheduled cycle time will remain constant on the route during the selected time period, we have the following:

$$c = n_0 h_{S_0} = n_1 h_{S_1} \tag{6.9}$$

Finally, we set $h_{E_1} = h_{S_0}$ and $\widehat{CV}_0 = \widehat{CV}_1$ and find n_1 in terms of n_0:

$$h_{E_1} = h_{S_0} \implies h_{S_0} = h_{S_1}(1+\widehat{CV}_1^2) \implies n_1 h_{S_1} = n_0 h_{S_1}(1+\widehat{CV}_0^2) \implies n_1 = n_0(1+\widehat{CV}_0^2) \tag{6.10}$$

Thus, in order to bring the effective headway on a route down to its scheduled headway, as measured during a particular time period, the number of buses operating on the route must be increased by a factor of $1 + \widehat{CV}_0^2$. The square of the coefficient of variation, then, represents the percent increase in buses on the route needed to provide the level of service indicated on the schedule, given no change in operations management and control. This is a worst-case scenario; in reality, adding buses to a route could be combined with other service improvements. The Additional Vehicles Required metric, as defined here, is most useful when comparing different service interventions, such as the holding and short-turning strategies explored in this research, with each other or with a baseline level representing current service. If a control strategy leads to a significant reduction in Additional Vehicles Required, this indicates that the vehicles on the route are being used more efficiently.

6.2.3 Percentage value vs. integer value

The calculations above result in a percent change in the number of buses required on a route. In reality, this will be translated into a number of buses. When examining a single route independent of any resource-sharing with other routes, the number of buses required is found by rounding $\frac{c}{H_S}$ up to the next higher integer:

$$n = \left\lceil \frac{c}{H_S} \right\rceil \tag{6.11}$$

The practice of interlining allows fractional values for n. Interlining is the practice of using a single vehicle on multiple routes over the course of a day. For example, consider two routes, each with a cycle time of 70 minutes and a target headway of 20 minutes. If the routes are scheduled independently, then $\lceil \frac{70}{20} \rceil = 4$ buses would be required on each, for a total of 8 buses. However, if they share a terminal and can be scheduled using the same vehicles, then interlining can be implemented by considering the two together as a combined route with a cycle time of 140 minutes, requiring only $\lceil \frac{140}{20} \rceil = 7$ buses. This can be thought of as 3.5 buses on each route.

Large agencies such as the MBTA, which operate many routes out of shared terminals, have many opportunities for interlining, including much more complex examples than the one given above, made possible by the use of optimization schemes. In the Fall 2014 bus schedule, which is the schedule used in this thesis, vehicles on Route 1 interline with 10 other routes. Vehicles on Route 28, on the other hand, do not interline except in special cases, as Route 28 draws from a separate fleet of articulated buses which are used on only a few MBTA routes.

Even when interlining is used, schedulers are limited to a discrete set of values for n, the number of buses used on a route. Without knowing the details of the specific schedule-optimization methods used, it is impossible to determine how many additional buses would be required for a particular scheduled headway. When presenting results for the AVR metric, we will give two values: The percentage change in n, which is an "ideal" value that could be achieved given perfect interlining, and the rounded value, which is the worst-case number of additional vehicles required if no interlining is possible.

6.3 Impact of even-headway strategy

The implementation of an even-headway strategy using an automated decision tool should improve bus performance in two ways. First, by adjusting departure times to even out headways, and second, by improving operator compliance with assigned departure times. As an even-headway strategy can only be implemented with some type of direct communication with the operator, whether through an automated device or through a supervisor or dispatcher, we expect that operator compliance will improve significantly. In this section we investigate the impacts of these two factors.

6.3.1 Improvement from reducing deviations

First, we examine the impact of reducing operator deviations from instructed departure times. We begin with the full values for deviations determined in Chapter 6, and then reduce those values by a factor of two. The original and reduced parameter values are given in Table 6.1. This represents a realistic reduction in deviations, as the halved values are similar in magnitude to those found by Milkovits (2008). The results are given in Table 6.2. The improvement in average passenger wait time from reducing the magnitude of deviations is non-linear; reducing from the full deviations to half deviations has a larger impact than going from the halved deviations to zero deviation. This non-linear relationship is consistent with the fact that wait time is a function of the square of the coefficient of variation of headways.

Table 6.1: Reduction in parameters for departure deviations (seconds of deviation)

	Full deviations		Half deviations	
	Avg Earliness	Avg Lateness	Avg Earliness	Avg Lateness
Dudley	123	147	62	74
Harvard	61	90	31	45
Mattapan	59	160	30	80
Ruggles	77	153	39	77

Table 6.2: Improvement in average passenger wait time from reducing deviations (using the schedule-following strategy)

Average passenger wait time	Route 1		Route 28	
(minutes)	Peak	Midday	Peak	Midday
Full deviations	7.26	8.64	6.14	6.54
Half deviations	7.07	8.39	5.97	6.28
Zero deviations	6.99	8.37	5.90	6.20

6.3.2 Improvement from implementing even-headway strategy

In this section, we examine the improvement from implementing the even-headway strategy. We use the halved deviations in both the even-headway and schedule-following versions of the simulation, in order to create an "apples-to-apples" comparison. The improvement from implementing the strategy is consistently larger than the improvement from halving the deviations, and the total reduction in APWT ranges from 0.68 to 0.76 minutes.

Table 6.3: Combined improvement from both halving deviations and implementing strategy

Average passenger wait time	Route 1		Route 28	
(minutes)	Peak	Midday	Peak	Midday
Schedule	7.07	8.39	5.97	6.28
Even-headway	6.55	7.88	5.39	5.87
Improvement from strategy	0.51	0.51	0.58	0.41
Improvement from deviations	0.19	0.25	0.17	0.27
Total improvement	0.70	0.76	0.75	0.68

To demonstrate the use of the Additional Vehicles Required metric, we show in Figure 6.4 the improvement over the three time periods defined above. In the AM Peak on Route 28, the largest improvement is observed, with the AVR dropping from 32.0% to 11.9% when the even-headway strategy is implemented, along with an improvement of 4.6% from reducing the deviations by half. This means that a total improvement of 24.7% was achieved, implying that implementing the even-headway strategy on this route during this time period had an effect equivalent to a 24.7% increase in the number of vehicles devoted to this route. The other two time periods, the AM Peak on Route 1 and the Midday School period on Route 28, show improvements of 7.7% and 7.4%, more modest in magnitude but still significant to an agency's bottom line.

It is also useful to examine the total amount of service provided in each scenario. Cer-

Table 6.4: Improvement in Additional Vehicles Required

	Route 1	Route 28	
	AM Peak	AM Peak	Midday School
Schedule	9.7% (1 bus)	32.0% (4 buses)	10.1% (2 buses)
Even-headway	6.4% (1 bus)	11.9% (2 buses)	7.4% (1 bus)
Improvement from strategy	3.4% (0 buses)	20.1% (2 buses)	2.7% (1 bus)
Improvement from deviations	4.3% (1 bus)	4.6% (0 buses)	4.7% (0 buses)
Total improvement	7.7% (1 bus)	24.7% (2 buses)	7.4% (1 bus)

tain types of even-headway strategies may appear to reduce average waits by evening out headways, but might in fact be simply increasing the number of trips made by reducing the average layover. This is unlikely to be the case with our strategy, because of the constraint preventing departures before the scheduled departure time. Nevertheless, for completeness we present in Table 6.5 the number of vehicle-timepoints served in each scenario; that is, the number of times each timepoint was served by a vehicle in each direction. This shows that in fact, slightly less total service was provided under the even-headway strategy, making the improvements in passenger waiting time even more impressive.

Table 6.5: Vehicle-timepoints served

	Route 1		Route 28	
	Peak	Midday	Peak	Midday
Schedule	535	592	736	1002
Even-headway	529	590	724	1009

6.4 Additional control strategies

In this section, we examine different strategies beyond the terminal-only holding strategy. These strategies rely on specific route characteristics, and so we limit our tests in this section to Route 1 only, which has more appropriate conditions for both strategies. First, we add to the terminal control points an additional mid-route control point in each direction, testing both a limited and unrestricted holding strategy. Next, we test a short-turn strategy that was identified during the experiment. In both cases, the new strategy is in addition to the terminal-based holding strategy, rather than a standalone strategy. This is because both midpoint holding and short-turning suffer from negative customer impacts that are not incurred by terminal-based holding, and so to minimize these impacts, the strategies are best used in combination with terminal holding.

6.4.1 Midpoint holding

The midpoint holding strategy tested here is conceptually similar to the basic even-headway strategy already discussed. In this strategy, in addition to controlling departures from the terminals, we select a single midpoint in each direction at which buses will be held to even headways. Providing additional control points should lead to better regularity of headways, and reduce the magnitude of holding necessary at each control point.

The cost to passengers of holding at a midpoint is greater than at a terminal, because at any midpoint stop, there are likely to be passengers on-board the vehicle who are being delayed in the middle of their trip. This is a very unpleasant experience for passengers, and was cited in conversations with MBTA staff as a reason the agency has avoided midpoint control strategies in the past. Pangilinan (2008) limited midpoint holding to a maximum of two minutes in his experiment with Chicago bus routes for this reason, and Sanchez-Martinez (2014) also uses a two-minute limit for midpoint holding. In our simulation, we test both unrestricted holding at the midpoint as well as holding limited to a maximum of two minutes.

We selected the Massachusetts Avenue subway station as the control point in both directions on Route 1. This stop is a good candidate for holding as it is one of the major transfer points on the route. Passengers who are connecting to another transit service would be very frustrated if the bus were held *before* the transfer point; conversely, passengers connecting to the Route 1 from the subway may be grateful that the bus was held at the transfer point if the hold allows them to make this connection. Figure 6-4 shows that the Mass Ave station (represented by a vertical black line) comes at a point where the number of passengers traveling through (neither boarding nor alighting) is relatively low (during many periods there is a "dip" in the load profile around this station), and the number of passengers boarding and alighting is relatively high. These two factors may help mitigate the in-vehicle holding time cost. Massachusetts Avenue was selected over the other major subway station, Hynes Convention Center, as it has bus pull-outs which would allow buses to hold without interrupting traffic on the street.

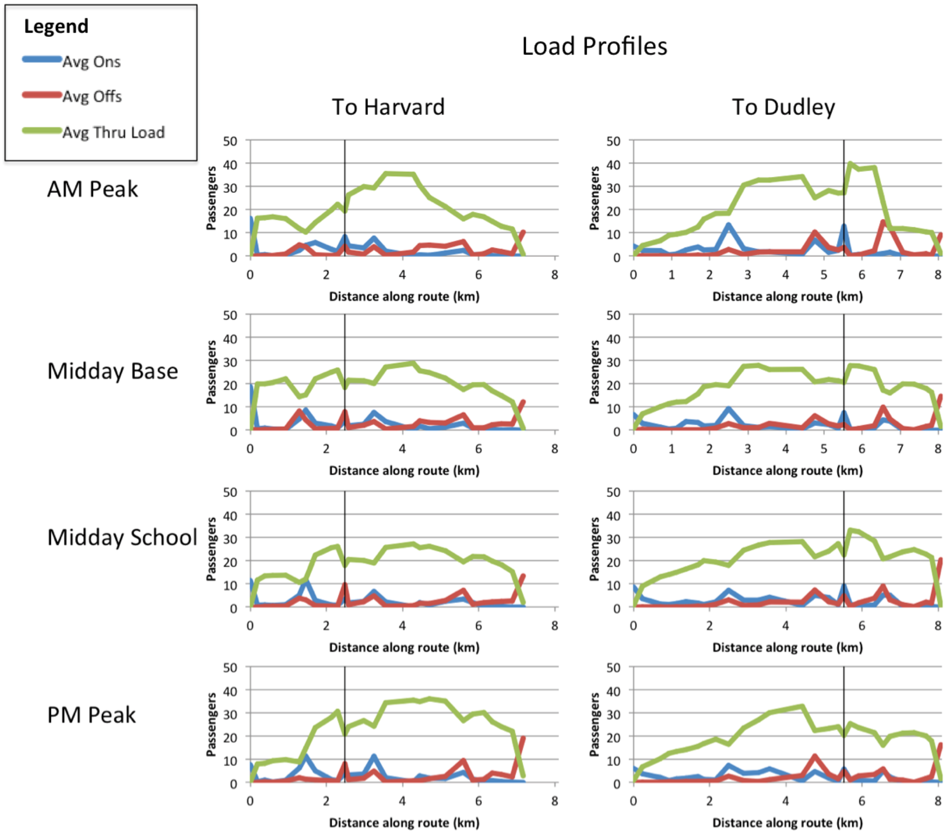

Figure 6-4: Load profiles on Route 1 by time period and direction

6.4.2 Short-turn

The final strategy simulated is the short-turn strategy described in Chapter 4. This strategy (shown in Figure 6-5) is implemented at the second-to-last stop approaching Harvard, at Bow Street and Massachusetts Avenue. A bus at this stop can turn down Bow Street and begin its next trip immediately, cutting off significant distance and travel time from the route. This short-turn strategy was noted by MBTA staff as a convenient option that is occasionally used by experienced dispatchers or supervisors. The added cost to passengers is that all of those on board a bus that is short-turned will be forced to alight, and either walk, or wait for the next bus. Mitigating this is the fact that the distance from the short-turn stop to the final stop is very short (approximately 850 feet) and that the short-turn strategy is likely to be implemented in cases where bunching is occurring, meaning that another bus will likely be available to pick up passengers quickly.

In this simulation, we will implement the strategy using a heuristic similar to that used for the holding strategies. The goal of the decision-rule is to use short-turning to improve

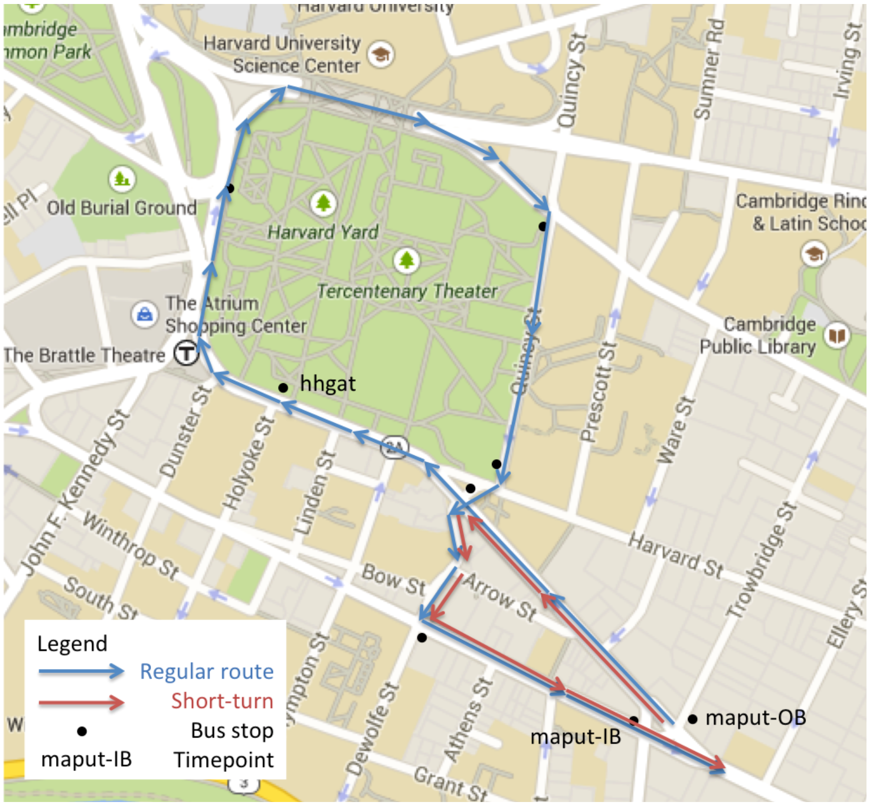

Figure 6-5: Short-turn strategy with timepoints marked

the evenness of headways; thus, it will be used when a vehicle at the control point has a long leading headway and a short trailing headway.

Figure 6-6 shows the basic structure of the route and the strategy, as modeled in the simulation. As the simulation is at the timepoint rather than the stop level, we approximate the location of the short-turn by its nearest timepoint, which is the timepoint at Mt. Auburn Street and Putnam Ave, abbreviated "maput." Each timepoint is treated separately in each direction, so the Mt. Auburn and Putnam timepoint is labeled as "maput-IB" and "maput-OB" in the "Inbound" (toward Dudley) and "Outbound" (toward Harvard) directions. The Harvard terminal and directional stops are represented here as simply "hhgat" for simplicity. The strategy is implemented as follows:

1. A bus arrives at maput-OB (this will be referred to as the "control vehicle").

2. If any buses are active on the route between maput-OB and maput-IB, do not short-turn.

3. If any buses are scheduled to enter service at Harvard before the control vehicle departs,

do not short-turn.

4. If neither of the above two conditions holds:

   (a) Predict the next departure time from maput-IB following the control vehicle, assuming that it does not hold beyond the scheduled departure or short-turn.

   (b) Calculate the ideal departure time for the control vehicle from maput-IB as the time that would equalize the vehicle's leading and trailing headways at maput-IB.

   (c) Predict the arrival time of the control vehicle at maput-IB if it follows the short-turn strategy and if it follows the regular route.

   (d) Choose the strategy which leads to an arrival time closest to the ideal departure time calculated in 4b.

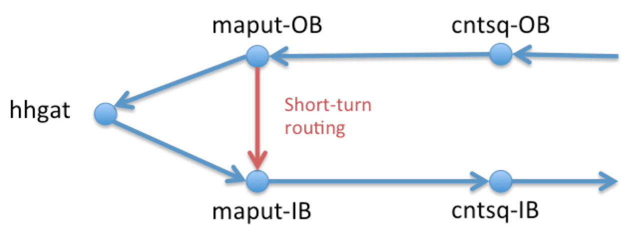

Figure 6-6: Implementation of the short-turn strategy in the simulation

To create a distribution of running-time data on the short-turn segment, we must use APC data, as the AVL data source is only available at the timepoint level. We create a set of short-turn running times by combining a vehicle's running time on the inbound segment from Mt. Auburn and Putnam with that vehicle's subsequent running times on the outbound segments between Quincy St and Mt. Auburn and Putnam. As the APC system is installed on a limited number of buses, the set of observed running times is smaller. However, because the mean running times on these segments do not vary greatly across time periods, we combine them into a single sample for the entire day, giving a total of 162 observations, with a mean of 176 seconds and a standard deviation of 45 seconds.

In addition to the running time on the short-turn segment, we must account for the additional dwell time necessary to unload all passengers on board the bus. As dwell time is not explicitly modeled in our simulation, we add a constant one minute to the running time to account for the additional dwell time necessary to instruct passengers and allow them to disembark.

6.5 Impact of midpoint holding

In this section, we examine the results of the midpoint-holding strategy, with and without a 2-minute limit on holding time, compared with the terminal-only holding strategy and the schedule-following strategy. The cost in time to passengers is decomposed into time spent waiting at stops and time spent holding on-board buses, and these are compared to determine the relative benefit of adding control points along the route.

Tables 6.6 and 6.7 show a comparison of the results from the terminal and midpoint strategies in the peak periods and midday periods, respectively. In these tables, "Wait time" refers to the waiting time of passengers at stops, while "In-vehicle holding" refers to the time spent by passengers in a vehicle holding at a midpoint. The tables show the savings in average passenger wait time, the cost in average minutes of in-vehicle holding time per passenger, and the average total delay which is simply the sum of the two sources of delay. All values represent the average per day over 100 simulations.

Table 6.6: Comparison of strategies in the peak periods on Route 1

	APWT (minutes)	Avg. in-vehicle holding time (minutes)	Avg. total delay time (minutes)
Schedule	7.07	0	7.07
Even-headway (terminals only)	6.55	0	6.55
Even-headway (terminals + midpoints)	5.99	0.60	6.59
Even-headway (terminals + midpoints, with 2-min limit)	6.28	0.25	6.53

Table 6.7: Comparison of strategies in the midday periods on Route 1

	APWT (minutes)	Avg. in-vehicle holding time (minutes)	Avg. total delay time (minutes)
Schedule	8.39	0	8.39
Even-headway (terminals only)	7.88	0	7.88
Even-headway (terminals + midpoints)	7.27	0.62	7.89
Even-headway (terminals + midpoints, with 2-min limit)	7.72	0.24	7.97

Table 6.8 shows the Additional Vehicles Required metric, as calculated for the different holding strategies. Although none of the values is sufficiently large to constitute an additional vehicle, the percentage change from 9.7% for the schedule-following strategy to 5.0% for limited midpoint holding constitutes almost half a bus (given the scheduled level of 10 buses in service), and thus could represent significant resources if interlining is possible.

Table 6.8: Additional Vehicles Required on Route 1 in the AM Peak period

	Additional Vehicles Required
Schedule	9.7% (1 bus)
Even-headway (terminals only)	6.4% (1 bus)
Even-headway (terminals + midpoints)	3.3% (1 bus)
Even-headway (terminals + midpoints, with 2-min limit)	5.0% (1 bus)

Overall, there appears to be no significant benefit to adding midpoint holding on Route 1. Holding vehicles at Mass Ave station did not reduce the average delay time to passengers, instead re-distributing a small portion of the delay from waiting time to in-vehicle holding time. In the next sections, we will discuss the relative importance of waiting time and in-vehicle holding time, and the change in the way delays were distributed across passengers.

6.5.1 Relative weights of wait time and holding time

The assessment of the costs of midpoint holding strategies depends upon the relative weights of wait time at stops and in-vehicle holding time. Significant research has been devoted to the relative disutility of different components of journey time. In this section, we will discuss some of this research, and how it applies to the question of in-vehicle holding time vs. waiting time.

Studies of travel behavior typically decompose travel time into two components: in-vehicle time and out-of-vehicle time, sometimes splitting out-of-vehicle time into access time (e.g. walking to a stop), initial wait time, and transfer time (Ben-Akiva and Lerman, 1985). Typically, wait time is found to be more onerous to passengers than in-vehicle time by a factor of 2 to 3 (Iseki et al, 2006). Sanchez-Martinez (2014) and Delgado et al (2012), in their simulations of holding strategies, used cost functions that included a factor of 2 for this value.

In the literature, in-vehicle time is typically treated as a single value, with no separate treatment for time spent holding in vehicles. The reason for this is simply a lack of data: In order to evaluate the impact of holding time on passengers, one would need detailed records of holding time on vehicles as well as specific data on individual passengers' trips by vehicle. Such data has only very recently become available, and only on the largest transit systems. In order to determine how to treat in-vehicle holding time, we will examine related literature.

Carrel et al (2013) examined the effects of various attributes of service on San Francisco Muni travelers' likelihood of shifting modes away from transit. They obtained two findings that are relevant to this research:

1. "[...] in-vehicle delays are more likely to drive people away from transit than longer waiting times at passengers' origin stops." This notably contradicts the research mentioned above.

2. Delays caused by traffic, vehicle breakdowns, and other causes easily visible to passengers had a significantly weaker effect than delays whose causes were not visible, such as congestion at tunnel portals.

The first finding may be related to the increasing availability of real-time vehicle arrival predictions. Authors such as Chow et al (2013) and Ferris et al (2010) have found that the availability of predicted arrival times significantly reduces passengers' estimation of their wait time and increases satisfaction with transit service. The second finding, while not specifically applied to holding strategies (which are not commonly used by Muni), likely applies to the case of holding. Even if a hold is announced to passengers, the reason for the hold is not obvious, and its benefits are entirely hypothetical, from the point of view of the passengers already on-board.

For these reasons, we believe that in-vehicle holding time is significantly more onerous to passengers than in-vehicle travel time, and should not be treated in the same way. Without a strong reason to believe that in-vehicle holding time has a stronger (or weaker) effect on passengers' utility than wait time, we treat the two equally, and sum them to create a measure of total delay time, which leads to the conclusion that total passenger cost is not improved by the addition of midpoint holding. We acknowledge that further research is needed in this area, perhaps using a targeted survey of the type implemented by Chow (2014), in combination with archived data from a decision tool, as described in Section 3.4.5.

6.5.2 Distribution of passenger cost

We have shown that, in terms of total passenger delay time, adding midpoints to the holding strategy has little effect; it simply trades in-vehicle holding time for waiting time, leaving the total delay approximately constant. Another way in which the strategy may be an improvement for passengers is if it reduces the variation in total passenger cost. Tables 6.9 and 6.10 show the median and 95th percentile values of passenger wait time and in-vehicle holding time.

Table 6.9: Distribution of passenger costs during the peak periods on Route 1

	Percentile values of waiting time (min)		Percentile values of in-vehicle holding time (min)	
	50th	95th	50th	95th
Schedule	5.6	18.7	0	0
Even-headway (terminals only)	5.3	16.7	0	0
Even-headway (terminals + midpoints)	5.0	15.1	0	4.4
Even-headway (terminals + midpoints, with 2-min limit)	5.1	15.9	0	2

It is clear that the variation in waiting time is reduced by the addition of midpoint holding. The limited midpoint holding strategy reduces the 95th percentile waiting time by 0.8 minutes in the peaks, and 0.3 minutes in the midday, likely because the 2-minute

Table 6.10: Distribution of passenger costs during the midday periods on Route 1

	Percentile values of waiting time (min)		Percentile values of in-vehicle holding time (min)	
	50th	95th	50th	95th
Schedule	7.0	20.9	0	0
Even-headway (terminals only)	6.8	18.9	0	0
Even-headway (terminals + midpoints)	6.4	17.2	0	4.6
Even-headway (terminals + midpoints, with 2-min limit)	6.6	18.6	0	2

constraint is more restrictive relative to the longer midday headways. Unfortunately, because the simulation does not model individual origin-destination pairs, we cannot create a distribution of total passenger delay time. However, if wait time and in-vehicle holding time are negatively correlated, then we can infer that the distributions of total delay time in the midpoint holding scenarios are still narrower than the distribution for terminal-only holding.

To show that wait time and in-vehicle holding time are negatively correlated, we first note that any passenger who arrives during a hold will, by definition, have zero wait time and positive in-vehicle holding time. For the remaining passengers, the negative correlation between wait time and in-vehicle holding time is created by the fact that vehicles are held when their leading headway is shorter than their trailing headway, and so on average, held vehicles have shorter leading headways than vehicles not held (this is shown in Figure 6-7). Passengers on-board a vehicle with a short leading headway most likely experienced a short wait time, and vice-versa. Therefore, for individual passengers, longer wait times are associated with shorter in-vehicle holding times.

Because of this inverse relationship between wait time and in-vehicle holding time, we can generally say that the midpoint holding strategies spread out the passenger cost more evenly among passengers, by reducing the wait time of the longest-waiting passengers, but adding holding time to those passengers who had short waits. The magnitude of the reduction in 95th-percentile total delay time is, at most, equal to the reduction in 95th percentile waiting time shown in Tables 6.9 and 6.10. The reductions of 1.6 minutes in the peak and 1.7 minutes in the midday periods, shown by the unlimited holding strategy, are significant, but due to the large amount of in-vehicle holding time necessary (greater than 4 minutes in the 95th percentile case), it is unlikely that these would be acceptable. In the more realistic scenario of midpoint holding with a 2-minute cap, the reduction in 95th-percentile wait time is much smaller, particularly in the midday as mentioned above.

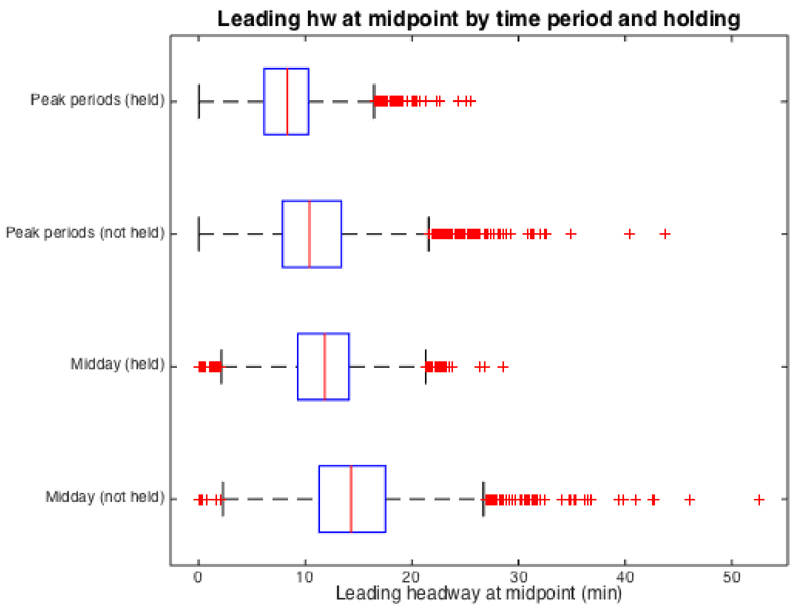

Figure 6-7: Leading headways by time period and holding strategy (Route 1)

6.6 Impact of short-turning

The results of the short-turning strategy are shown in Tables 6.11 and 6.12. The improvement in average passenger wait time, and the distinction between the peak and midday periods, as well as the costs in forced alightings, will be discussed in detail in this section.

Table 6.11: Impact of short-turn strategy in the peak periods (at Harvard, Route 1)

	APWT (minutes)	Percentile values of waiting time (min)		% of trips short-turned	% of outbound passengers forced to alight
		50th	95th		
Schedule	7.07	5.6	18.7	-	-
Even-headway (terminals only)	6.55	5.3	16.7	-	-
Even-headway (terminals + short-turn)	6.33	5.2	15.9	2.8%	2.9%

Table 6.12: Impact of short-turn strategy in the midday periods (at Harvard, Route 1)

	APWT (minutes)	Percentile values of waiting time (min)		% of trips short-turned	% of outbound passengers forced to alight
		50th	95th		
Schedule	8.39	7.0	20.9	-	-
Even-headway (terminals only)	7.88	6.8	18.9	-	-
Even-headway (terminals + short-turn)	7.84	6.8	18.7	5.7%	3.2%

6.6.1 Improvement in APWT

The results of the short-turn strategy are strikingly different between the peak and midday periods. In the peaks, the short-turn strategy reduced average passenger wait time by an additional 0.22 minutes over terminal-only holding, and reduced the 95th percentile wait time by an additional 0.8 minutes. In the midday, on the other hand, the short-turn strategy only saved 0.04 additional minutes of APWT and reduced the 95th-percentile wait time by 0.2 minutes.

To understand this difference, we examine the wait-time tradeoff between passengers waiting on the skipped segment, who face a cost in additional wait time whenever the short-turn is implemented, and passengers waiting downstream, who benefit from reduced waiting time. We refer to the set of stops which are skipped by short-turned vehicles as the "Harvard Loop."

Figure 6-8 shows the cost to passengers on the Harvard Loop, in the form of distributions of APWT values over the 100 replications of the simulation. Figure 6-9 shows the benefits, in the form of reduced APWT on the remainder of the route. The difference between the two time periods shows up both on the cost and benefit side: As summarized in Table 6.6.1, the costs are greater, and the benefits smaller, in the midday than in the peaks. In the next section we will explain the mechanism behind this difference by examining the distributions of headways among short-turned vehicles.

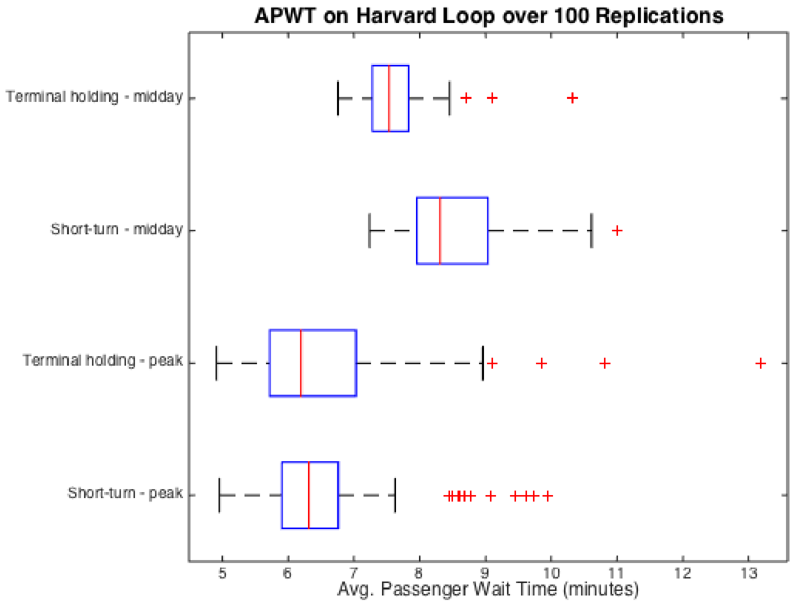

Figure 6-8: Distributions of average passenger wait time on the Harvard Loop

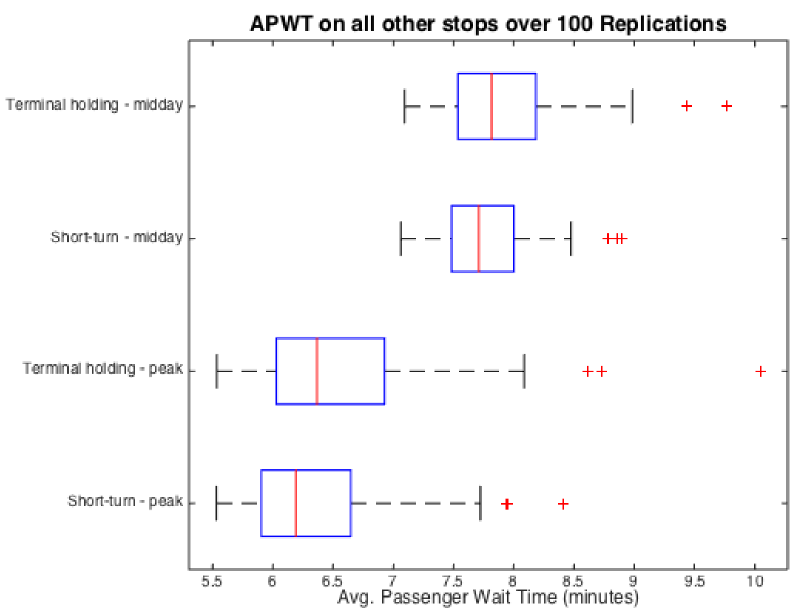

Figure 6-9: Distributions of average passenger wait time at all other stops

Table 6.13: Costs and benefits in TPWT from short-turn strategy in different periods

All values are means over 100 replications	Cost in add'l TPWT on Harvard loop (pass.-hrs)	TPWT savings at other stops (pass.-hrs)
Midday periods	4	5
Peak periods	0	13

6.6.2 Usage of short-turn in different time periods

As described above, the short-turn strategy had very different impacts on passenger wait time in the midday periods vs. the peak periods. Another difference between the two periods was in the frequency of short-turning; although the percentage of passengers forced to alight was approximately the same (2.9% in the peak vs. 3.2% in the midday), the percentage of trips short-turned was much higher in the midday (5.7% vs. 2.8%).

Figure 6-10 shows that the minimum leading headway required to justify a short-turn is approximately 10 minutes in all periods. 10-minute headways naturally occur more frequently during the midday, due to the presence of fewer vehicles in service, which is why the frequency of short-turning vehicles is higher during the midday. The reason the percentage of passengers forced to alight does not similarly rise in the midday is that the vehicles short-turned are on average less crowded relative to other vehicles in the same period; a 10-minute headway may even be shorter than the scheduled headway during the midday.

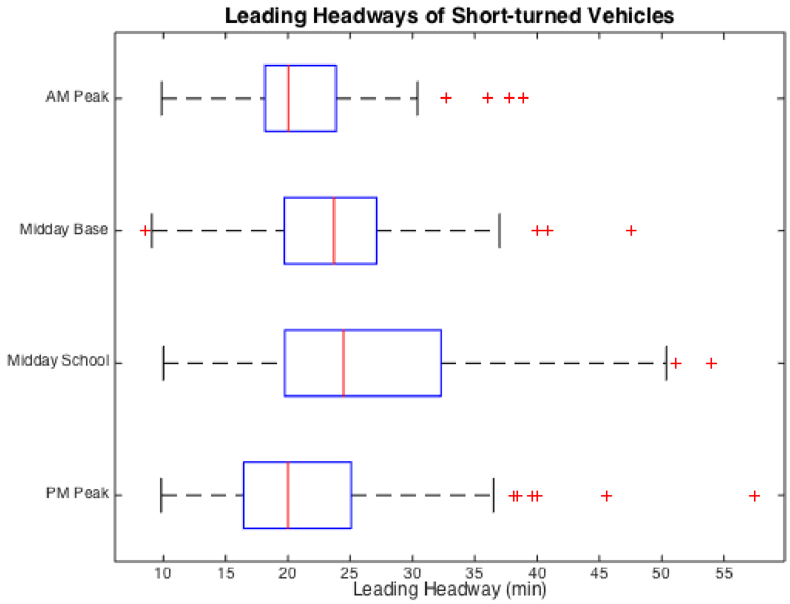

Figure 6-10: Leading headways of short-turned vehicles at the short-turn point

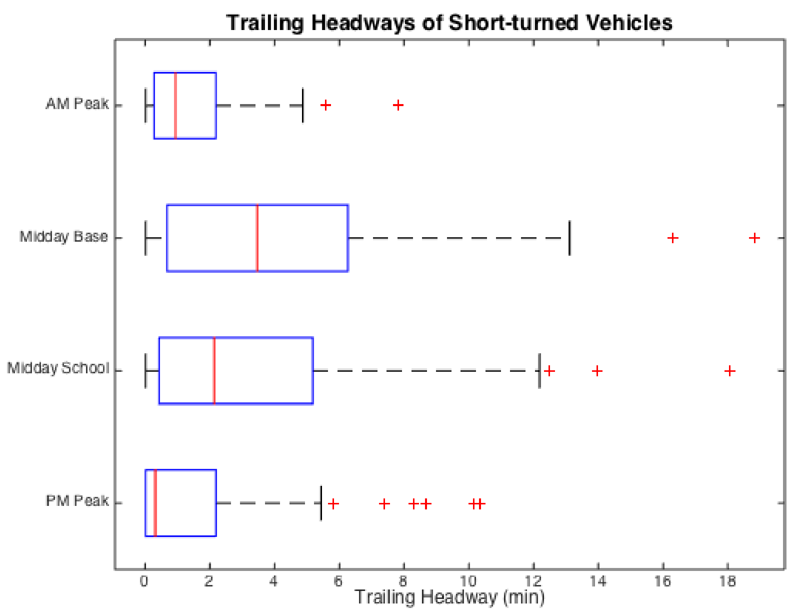

Figure 6-11: Trailing headways of short-turned vehicles at the short-turn point

One possible reason for the disparity might be that we need to apply a higher threshold to the short-turn decision in the midday periods than we do in the peaks. To test this, we extend the decision rule as follows: Given the ideal departure time from maput-IB t_{ideal}, the expected departure time with short-turning t_{short}, and the expected departure time without short-turning t_{normal}, we short-turn a bus at maput-OB if $|t_{ideal} - t_{short}| + d < |t_{ideal} - t_{normal}|$, for some threshold value d. The threshold value can be raised to decrease the amount of short-turning, and use it only in cases where the benefit to headway regularity is highest. With a value of zero, the strategy is identical to that describe above. We show the results with various threshold values in Table 6.14. The results are similar regardless of threshold, which confirms that the short-turning strategy provides no benefit over the terminal-only even-headway strategy in the midday periods. Reasons for this will be discussed in Section 6.6.3.

Table 6.14: Impact of short-turn strategy in the midday with varying threshholds (at Harvard, Route 1)

Threshold (minutes)	APWT (minutes)	Percentile values of waiting time (min)		% of trips short-turned	% of outbound passengers forced to alight
		50th	95th		
0	7.84	6.8	18.7	5.7%	1.6%
2	7.84	6.8	18.7	4.1%	1.3%
4	7.89	6.8	18.9	2.3%	0.8%
6	7.86	6.8	18.8	0.2%	0.0%

6.6.3 Impacts on groups of passengers

To explain the differences in performance of the strategy during different time periods, we examine the impact of a short-turn on passenger wait time. This analysis uses assumptions similar to those used by Eberlein (1999) to model deadheading and expressing. The assumptions are the following:

- Headways remain constant along the route
- Passenger arrival rates are constant
- The control vehicle will not hold at the terminal if it is not short-turned
- The trailing vehicle will neither hold at the terminal nor be short-turned

These assumptions reduce the realism of the model, but it is nevertheless illustrative of the relationships between the main factors influencing the costs and benefits of the short-turn. We will use the following notation for these factors:

λ_i = passenger arrival rate at stop i

S = the set of skipped stops (the Harvard Loop)

D = the set of downstream stops

h_l = leading headway of control vehicle

h_t = trailing headway of control vehicle

t_0 = running-time savings of short-turn

There are three groups of passengers who experience a change in wait time when the Bow Street short-turn is implemented:

Costs on the Harvard Loop

The first group consists of those passengers who would have boarded the control vehicle if it was not short-turned. These passengers arrived between the leading vehicle and the time that the control vehicle would have arrived, a length of time equal to the leading headway h_l. Thus, the number of passengers in this group is equal to $h_l \sum_{i \in S} \lambda_i$. The cost to each passenger is the additional wait time they face before boarding the trailing vehicle, which is equal to the trailing headway h_t. The total cost in passenger-minutes, then, is $h_t h_l \sum_{i \in S} \lambda_i$.

Costs to downstream passengers

The second group consists of those passengers downstream who would have boarded the control vehicle if it was not short-turned, but because of the short-turn, arrive too late and must wait for the trailing vehicle. They arrive on an interval of length t_0, the time between when the control vehicle would have arrived and when the trailing vehicle actually does arrive, so the total number of passengers in this group is $t_0 \sum_{i \in D} \lambda_i$. The cost to them is equal to their additional waiting time, which is h_t, and the total cost associated with this group is $h_t t_0 \sum_{i \in D} \lambda_i$.

Benefits to downstream passengers

The third group consists of those passengers downstream who board the control vehicle whether or not it is short-turned. These passengers arrive between the leading vehicle's arrival and the control vehicle's actual arrival (after implementing the short-turn), a time period of length $h_l - t_0$. The total number of passengers is $(h_l - t_0) \sum_{i \in D} \lambda_i$, and the benefit to these passengers is the reduction in waiting time, equal to t_0. The total benefit in passenger-minutes is $t_0(h_l - t_0) \sum_{i \in D} \lambda_i$.

Net benefit in passenger wait time

Based on the three groups outlined above, the net benefit NB of short-turning is estimated as shown in Equation 6.12, and the total number of affected passengers P as in Equation 6.13. The net wait time benefit per affected passenger is $\frac{NB}{P}$.

$$NB = -h_t h_l \sum_{i \in S} \lambda_i - h_t t_0 \sum_{i \in D} \lambda_i + t_0(h_l - t_0) \sum_{i \in D} \lambda_i \tag{6.12}$$

$$P = h_l \sum_{i \in S} \lambda_i + t_0 \sum_{i \in D} \lambda_i + (h_l - t_0) \sum_{i \in D} \lambda_i \tag{6.13}$$

This equation shows the basic mechanism by which the midday period differs from the peak periods. The mean values for the inputs to this equation are shown in Table 6.15. Between the two periods, the demand levels are approximately 55% higher in the peaks than the midday, an increase which is roughly constant between the loop segment and the downstream segment. The major differences are that headways are longer in the midday, because fewere buses are in service, and the time savings from the short-turn strategy is reduced, because running times on the loop are shorter due to reduced traffic. These differences lead to a increase in the magnitude of the first term (costs to passengers on the loop), and changes in magnitude of the other two terms which are unclear, but are dominated by the first term as headways increase. A numerical example is shown in Table 6.16, which uses the values from Table 6.15 as inputs to calculate the benefits and costs of each component and the net benefit.

These results show that the net benefit of the short-turning strategy did not simply vary with the level of demand on the route during the different periods, but that in fact the costs

Table 6.15: Example inputs to net benefit of short-turn (mean values from simulation)

	Leading headway h_l	Trailing headway h_t	Arrival rate on loop $\sum_{i \in S} \lambda_i$	Arrival rate downstream $\sum_{i \in D} \lambda_i$	Time savings t_0
Midday	25.0 min.	3.5 min	0.9 pass./min	4.0 pass./min	4.7 min
Peak	22.3 min.	1.2 min	1.4 pass./min	6.2 pass./min	5.7 min.

Table 6.16: Example calculation of net benefit of short-turn

	Cost on Loop per affected pass.	Downstream cost per affected pass.	Downstream benefit per affected pass.	Net benefit per affected pass.
Midday	-0.2 pass.-min.	-0.3 pass.-min.	3.5 pass.-min.	3.0 min.
Peak	-0.6 pass.-min.	-0.5 pass.-min.	3.1 pass.-min.	1.9 min.

were lower, and the benefits higher, during the peak. This simplified model still shows a positive net benefit in the midday example scenario, but the net benefit in both scenarios is overestimated due to the assumption of deterministic headways which ignores the loss from irregular running times along the route.

This demonstration shows that the benefits of short-turning can vary dramatically across time periods, and different characteristics such as demand patterns, headways, and running-time savings. If a heuristic strategy like this one is to be used on a route, significant simulation work should be done first to ensure that the strategy will be effective. To ensure that all factors are fully taken into account, an optimization-based approach would be ideal, with estimates (or real-time measures if available) of passenger loads and passenger demand levels as inputs in addition to running-time estimates.

It is also important to note that unusual circumstances could create specific cases in which the short-turning strategy performs poorly. For example, an incident on the Red Line subway service could lead to a high number of passengers boarding Route 1 at Harvard. This would result in a temporary increase in demand for service on the Harvard Loop, and thus an increase in the cost of short-turning. Without a way to measure the demand at a stop in real-time, it is impossible to adjust the strategy to react in real-time to such disruptions. This means that, in the absence of real-time data on demand at bus stops, a short-turning strategy like this one must be implemented in the context of a decision-*support* tool rather than a fully-automated tool, so that a dispatcher or supervisor makes the final decision to short-turn.

6.6.4 Cases of short trailing headways

Another reason to consider the trailing headways of vehicles that were short-turned is to track cases in which a short-turned vehicle will have a very short trailing headway. Such cases are noteworthy for two reasons:

1. Our simulation does not allow passing, and has every passenger board the first available vehicle. Therefore, it probably overestimates loads on the lead vehicle of a closely-bunched pair.

2. An alternative strategy for a closely-spaced pair of vehicles could be to short-turn the *trailing* vehicle, which likely has fewer passengers on board. Our heuristic rule does not handle this type of strategy, but it could be implemented in a real-world scenario, especially in a case where real-time data on passenger loads is available.

For both peak periods, the 75th percentile trailing headway is approximately 2 minutes, meaning that most short-turned vehicles have very short trailing headways. The midday periods have a much wider range of trailing headways, meaning that the strategy is used in a wider range of scenarios in the midday. This means that we are likely overestimating the number of passengers forced to alight in the peak periods, and also that an alternative strategy based on short-turning the vehicle with the smaller load would probably perform better in this period.

6.7 Summary

In this chapter, we have examined the impacts of reducing schedule deviations and implementing midpoint-holding and short-turning strategies. We defined metrics including passenger wait time and effective headway. We also created a new measure, Additional Vehicles Required, which uses the effective-headway concept to estimate how many additional vehicles would need to be added to a route, given a fixed level of variation of headways, to bring the effective headway down to the level of the scheduled headway.

The largest impact was found in the implementation of the terminal-only holding strategy, which was found to reduce the Additional Vehicles Required by an estimated 2 buses in the AM Peak period on Route 28, and 1 bus in the Midday School period on Route 28. Reducing deviations from assigned departure times by half the current average values (through improved training and supervision) would also improve service significantly, reducing the Additional Vehicles Required by 1 bus on Route 1 in the AM Peak period.

Adding mid-route holding points did not significantly improve the overall performance of the strategy. The midpoint strategy reduced waiting time at the cost of increased in-vehicle holding time, with this trade-off being approximately equal in passenger delay time. Whether this constitutes an improvement depends on how in-vehicle holding time is valued relative to waiting time at stops; based on existing research we suggest that in-vehicle holding time may be just as onerous as out-of-vehicle waiting time, but with no specific research in this area, it is unclear how to weigh the two forms of delay. One minor area of improvement is in the reduction in variation of total passenger cost; essentially, the midpoint holding "spreads out the pain" of delays among passengers.

The short-turn strategy was found to reduce passenger wait time significantly in the peak periods, but not in the midday periods. This was explained by the longer headways and reduced time savings in the midday period vs. the peaks, and the way in which these factors interact. The complexity of these relationships suggests that an optimization framework, rather than a heuristic approach, would be the ideal method to use for a decision-support tool. Also, since the short-turn impacts fell disproportionately on more-crowded buses during the peak, a more sophisticated version of the strategy that allows for short-turning the trailing vehicle of a bunched pair would probably be more effective. We emphasize the advisory (rather than compulsory) nature of short-turn suggestions from a decision-support tool, as unusual situations such as unexpected spikes in demand may quickly invalidate the tool's suggestions.

Overall, these results suggest that implementing a terminal-only holding strategy, along with improving operator compliance with instructions, should be top priorities for the MBTA. In addition, using the short-turn strategy during the peak periods would add a significant additional benefit, but a more sophisticated optimization should be developed that takes into account passenger load.

Chapter 7

Conclusion

This thesis has explored the uses of real-time data to improve bus service, using both experimentation and simulation. In this chapter, we first summarize the results of the previous chapters. We then make specific recommendations for the MBTA regarding its potential use of real-time data for operations management. Finally, we outline areas for future research.

7.1 Summary

This research began with a review of existing literature on transit service reliability, control strategies used to improve reliability, and experiments and simulations. This was followed by an investigation of real-time decision-support tools, in which a software application was designed and implemented using real-time data from MBTA buses. Using this application, an experiment was conducted, the results of which led to an investigation of issues with regulating terminal departures. To augment the experimental results, a simulation model of two MBTA bus routes was developed and validated. Finally, the simulation was used to test the effect of reducing operator deviations from assigned departure times, as well as two additional control strategies. The results are summarized in this section.

7.1.1 Literature review

Chapter 2 presented a review of existing research on factors affecting bus service reliability, control str. Abkowitz et al, Levinson, and Pangilinan describe measures taken to improve bus service reliability, including priority (signal priority or exclusive lanes), control strategies, and operational strategies (timetables and fleet management). They find that effective supervision and good communication are key to precise control of bus departure times.

As control strategies are an essential component of this thesis, we described the variety of control strategies that have been explored in the literature. Specific strategies include heuristics such as Turnquist's Prefol strategy and Bartholdi and Eisenstein's method of "self-equalizing headways", as well as optimization routines such as those developed by Delgado et al and Sanchez-Martinez. In some cases they have been tested experimentally, as in the work of Strathman et al, Pangilinan, and Xuan et al. In other cases, they are tested with simulations. Sanchez-Martinez developed the particular simulation framework which was used in this work.

7.1.2 Application design

In Chapter 3 we outlined the components of an automated decision-support tool, the decisions that must be made in implementing such a tool, and its various uses. We described a variety of control strategies that might be implemented using a decision-support tool, including holding, deadheading, expressing, and short-turning. These range in complexity and in data requirements, ranging from the simplest headway-based holding strategy which requires only the predicted arrival time of the trailing vehicle, to short-turning strategies which ideally would use real-time data on passenger loads.

The benefits of a decision tool include not only the implementation of control strategies, but also the archived data it produces, which can be used to improve performance analysis and service planning. When control strategies are applied in an ad-hoc way through radio or in-person communication, typically no records are made of the actual control decision, and so users of archived data must infer whether or not a control action was taken. Archived data from an automated decision tool allows much more precise knowledge of whether or not a control action was recommended, which then allows for more accurate tracking of on-time departures, dwell times, and run times.

A variety of data sources, decision algorithms, and user-interface options exist, but all must fit into a basic structure which we describe. The software components must download the necessary data, applying a decision algorithm, output the suggested control actions to users, and archive records of the control actions. Typically a "prediction interpreter" component will be needed to convert customer-facing data sources, which focus on predicted arrival times in the near future, into a useful format for decision tools, which often require a recent history of departure times.

Using this basic design, we developed a specific software application to provide decision support for a holding strategy to be implemented on MBTA Route 1 by supervisors using mobile phones. The application is specific to the MBTA context, but the decisions made in its implementation are common to any implementation of an automated decision-support tool. The process included selecting the data sources, method of delivering instructions, design of the user interface, and the structure of the archived data. The process is instructive for any large North American transit agency, many of which have similar IT infrastructure and face similar challenges as the MBTA.

7.1.3 Experiment

Using the decision-support tool described above, we performed an experiment on MBTA Route 1, a major crosstown route. The strategy consisted of a variant of Turnquist's Prefol strategy, a simple method of holding vehicles to equalize the leading and trailing headways. It was implemented at the route's two terminals, Harvard and Dudley Stations. The implementation of the strategy resulted in a reduction of approximately one minute of average passenger wait time in the Harvard-to-Dudley direction, but no improvement in the Dudley-to-Harvard direction.

Poor adherence to the suggested departure times provided by the mobile app was a major problem at both terminals, and significantly worse at Dudley. Deviations from suggested departure times were significantly more frequent and of a larger magnitude Dudley, and for those departures that were suggested to be made "ASAP", departures from Dudley were frequently delayed by as much as eight minutes. The causes of these deviations were identified as operator discipline, the layout and use of layover and passenger-boarding areas at Dudley, and long boarding times due to cash fare payment or adding value to CharlieCards on-board buses.

In addition to the results during the experiment, the scheduling on Route 1 was examined further. It was found that scheduled half-cycle times are typically shorter than the 95th percentile of running times in the Dudley-to-Harvard direction, another factor that can lead to irregular departures. In the peak periods, enough extra half-cycle time exists in the opposite direction to compensate for this by shifting the schedule, but in the midday there is simply insufficient total cycle time.

Based on the problems identified during the experiment, we made specific recommendations for operational improvements on Route 1, some of which are applicable across all MBTA bus routes:

- Better training of both operators and supervisors with respect to the importance of on-time terminal departures
- Management intervention with operators who show a pattern of poor departure-time discipline, using AVL-based reports
- Use of LED signs to display departure times at bus berths
- Use of the Route 1 berth at Dudley for layovers to allow passengers to board early
- Increased use of a handheld CharlieCard validator to allow boarding at the rear door on buses at Dudley
- Specific instructions to inspectors on when and how to use express, deadhead, or short-turn strategies

The results of the experiment highlighted the importance of precise control of terminal departure times, a result that echoes past work by Pangilinan (2006), Cham (2006), and Milkovits (2008). The importance of first improving terminal departures before implementing any control strategies will be one of the key recommendations of this thesis.

7.1.4 Simulation

In Chapter 5, we drew upon the work of Sanchez-Martinez (2014), who created a simulation model of a high-frequency bus route based on automatically-collected data. This model simulates running times at the timepoint level by drawing from a distribution of observed running times. Beginning with code written by Sanchez-Martinez for his PhD dissertation, we first modified the software to use MBTA data sources. We then added two elements that improve the validity of the simulation: A bivariate distribution of running times, and a detailed treatment of terminal departure behavior.

The bivariate distribution of running times is a concept described in earlier work by Sanchez-Martinez (2012), and implemented in our simulation. It extends the simple distribution of running times based on time of day, by using the running time of the same vehicle on the previous segment. This allows the correlation between successive running times on a single trip to be modeled, which is a key component of the mechanism that leads to bus bunching.

We used the work of Milkovits (2008) as a basis for our treatment of terminal departure behavior. We calculated recovery times at terminals as the minimum of two values: minimum required recovery time and available recovery time. The minimum required recovery time represents the amount of time needed by the bus operator to board and alight passengers and perform any other necessary tasks, and is modeled with a normal distribution. The available recovery time includes a random deviation component, which represents voluntary deviations from the assigned departure time on the part of the operator. The negative exponential distribution suggested by Milkovits was found to be a good fit for these deviations.

Using the bivariate distribution of running times combined with the terminal-departure model, we were able to enhance the original Sanchez-Martinez simulation code to accurately simulate MBTA Routes 1 and 28. Comparing the means and standard deviations of end-to-end running times and headways in the simulation against real-world observations showed that these enhancements improved the validity of the model over what had originally been achieved by Sanchez-Martinez with Transport for London.

7.1.5 Simulated experiments

In Chapter 6, we used the simulation model from Chapter 5 to examine the impacts of reducing schedule deviations and implementing midpoint-holding and short-turning strategies. We defined metrics including passenger wait time and effective headway. We also created a new measure, Additional Vehicles Required, which uses the effective-headway concept to estimate how many additional vehicles would need to be added to a route, given a fixed level of variation of headways, to bring the effective headway down to the level of the scheduled headway.

The largest impact was found in the implementation of the terminal-only holding strategy, which was found to reduce the Additional Vehicles Required by an estimated 2 buses in the AM Peak period on Route 28, and 1 bus in the Midday School period on Route 28. Reducing deviations from assigned departure times by half the current average values (through improved training and supervision) would also improve service significantly, reducing the Additional Vehicles Required by 1 bus on Route 1 in the AM Peak period.

Adding mid-route holding points did not significantly improve the overall performance of Route 1. The midpoint strategy reduced waiting time at the cost of increased in-vehicle holding time, with this trade-off being approximately equal in passenger delay time. Whether this constitutes an improvement depends on how in-vehicle holding time is valued relative to waiting time at stops; based on existing research we suggest that in-vehicle holding time may be just as onerous as out-of-vehicle waiting time, but with no specific research in this area, it is unclear how to weigh the two forms of delay. One area of improvement is in the reduction in variation of total passenger cost; essentially, the midpoint holding "spreads out the pain" of delays among passengers.

The short-turn strategy was found to reduce passenger wait time significantly in the peak periods, but not in the midday periods. We showed that the reason for this difference is that, given a roughly constant value of time savings from the short-turn strategy, an increase in headways leads to a large increase in costs and a smaller increase in benefits of short-turning. In addition, since the short-turn impacts fell disproportionately on more-crowded buses during the peak, a more sophisticated version of the strategy that allows for short-turning the trailing vehicle of a bunched pair would probably be more effective.

Overall, these results suggest that implementing a terminal-only holding strategy, along with improving operator compliance with instructions, should be top priorities for the MBTA. In addition, the specific short-turn strategy that was tested was found to add a significant additional benefit, but in order to implement such a strategy, an optimization should be developed that takes into account real-time or estimated passenger loads.

7.2 Recommendations

The MBTA should move towards a more systematic use of real-time data for operations control and performance management. Specific recommendations were outlined in Chapters 4 and 6, but in general, they fall into three categories: improvements to terminal departure adherence, improved uses of existing technology for operations management and planning, and the implementation of automated control strategies.

7.2.1 Terminal departure adherence

This research has shown that on-time terminal departures are a key driver of reliability throughout a route, and a major factor in explaining poor performance on the MBTA bus routes studied. To combat poor schedule adherence at terminals, the MBTA should institute a policy of tracking on-time departure performance by individual bus operators at terminals. This could be used, as it has been at agencies in Denver and Minneapolis, in regular counseling meetings between garage managers and the worst-performing bus operators. As emphasized in Chapter 4, these improvements to departure discipline should be made *before* any type of systematic control strategy is implemented.

The MBTA should also examine terminal-specific factors that influence departure times, including the allocation of space for layovers, the location of operator break facilities, and the locations of fare vending machines. These physical factors, as discussed in Chapter 4, can have a strong effect on the ability of bus operators to depart on time.

Finally, the MBTA should examine factors influencing dwell time, including payment with cash and adding value to CharlieCards on-board the bus. The long dwell times associated with passengers adding value to CharlieCards are a result of the MBTA's fare structure and the lengthy process of paying in cash or adding value on a bus. Both the fare policy and the technology used to accept payment should be re-examined with a focus on dwell-time impacts. In addition, the use of hand-held CharlieCard readers to speed up boarding times should be implemented at major stops such as Dudley which suffer from long dwell times.

7.2.2 Operations management and planning

MBTA supervisors currently use the technology available to them in a very limited way. In order to utilize supervisors more effectively, the MBTA should develop applications aimed at the information needed by supervisors in the field, including, for example, information on vehicle and crew schedules. Improved software can also enhance the effectiveness of dispatchers; for example, exception-based reporting that highlights early departures from terminals when they happen can be used to provide real-time feedback to operators on their performance.

Operations planning is another area where automatically-collected data should be used to enhance service. The MBTA should analyze running times from AVL data more frequently to ensure that sufficient half-cycle time is provided in the schedule. Also, the Additional Vehicles Required metric developed in this thesis can be used to identify routes where resources are being used inefficiently due to high variability in service, and where targeted control interventions or other service improvements could lead to a reduction in resources required.

7.2.3 Automated control strategies

The MBTA should move toward implementation of automated control strategies on its routes. Off-the-shelf products exist that allow for implementation of holding strategies on bus, light-rail, or heavy-rail routes. Rail lines are good candidates for trial runs, as they have dedicated dispatchers who could observe automated decisions and override them if anything went wrong. The Mattapan Trolley line in particular is frequently used to test technologies for later use in other parts of the system. On bus routes, automated control strategies could be implemented with devices installed in buses, or through the use of electronic signs such as the ones installed at Dudley, Ruggles, and other major bus terminals, which could be used to display dynamic departure times.

At a minimum, a policy of holding at terminals to improve headways should be applied. It is clear from this research, as well as past work, that holding at terminals can lead to significant improvements in passenger wait times. Holding at midpoints was shown not to provide an additional benefit over terminal-only holding, and so terminals should be the primary focus for interventions. The short-turn variant used in the simulation model of Chapter 6 should also be explored, although a more sophisticated model should be used in order to more precisely account for the costs and benefits.

7.3 Future research

During the course of this research, many opportunities for future research have been identified. In this section, we will describe areas with potential in the use of real-time data and mobile technology to improve transit service. The three main areas identified are software frameworks, simulations, and the passenger experience of holding strategies.

7.3.1 Software frameworks

In this thesis, a framework for a decision-support tool using real-time data was developed and used to create a mobile app supporting a control strategy. In addition, a simulation framework developed in previous research was extended to model a pair of MBTA bus routes. Future work on software for transit research could develop a platform to combine these two types of software, by specifying an API for simulation software to communicate with decision-tool software. Using such an API, researchers could easily test different control strategies in simulations as well as in actual transit operations.

7.3.2 Simulations

With the increasing availability of automatically-collected data, and recent advances in analyzing these data such as origin-destination inference, many opportunities exist to improve upon the simulation used in this research.

- The simulation in this thesis used timepoint-level running-time data, due to the limitations of the AVL system. Additional data sources such as the Automated Stop Announcement system could be used to infer stop-level arrival and departure times.
- Our simulation used arrival rates and alighting fractions to model the loads on board each bus. Using origin-destination inference, scaled appropriately using Automated Passenger Counter data, individual passenger trips could be modeled explicitly, allowing for disaggregate measures of passenger wait time, in-vehicle holding time, and total delay.
- Dwell time was not explicitly modeled in our simulation, although dwell-time effects represent an important component of the benefits of holding strategies. The dwell-time model used by Milkovits (2008) could be re-examined using the more-accurate AVL and AFC data available today, to further refine the model.

In addition to these specific improvements, our investigation of a specific short-turn strategy suggested that an optimization-based approach would be best-suited to short-turning and other strategies. The optimization framework developed by Sanchez-Martinez (2014) could be expanded to include the possibility of short-turning. Since short-turn options are typically limited to a few convenient locations on a route, and their running-time impacts are fixed for a particular time of day, adding short-turning possibilities to the set of holding options used by Sanchez-Martinez should not add significant computational complexity to the problem.

7.3.3 Passenger experience

As discussed in Chapter 6, there is a significant need for further research in the area of passenger experience of in-vehicle holding time. Although we have recommended against implementing midpoint holding on bus services, midpoint holding is often necessary on subway and light rail, as delays propagate down the line. As increasing numbers of agencies use automated tools to support and track holding strategies, detailed datasets on holding times will become available. These can be used in a framework such as a discrete-choice model to estimate the disutility of in-vehicle holding time for passengers, and allow agencies to make well-informed decisions about the tradeoffs of midpoint holding. In addition to discrete-choice modeling, the prompted-recall survey methodology developed by Chow et al (2014) could be used to survey passengers and match their responses with automatically-collected data on in-vehicle holding time.

Appendix A

Headway Ratio Analysis

Table A.1: Headway Ratio at Inbound Time Points by Headway Ratio at Harvard

Headway Ratio at Harvard	Total Trips	Harvard			Hynes			Dudley		
		Mean	Std. Dev.	Coeff. of Var.	Mean	Std. Dev.	Coeff. of Var.	Mean	Std. Dev.	Coeff. of Var.
0 to 0.4	100	0.22	0.12	0.53	0.43	0.46	1.07	0.67	0.72	1.08
0.4 to 0.8	157	0.65	0.11	0.17	0.63	0.41	0.65	0.66	0.54	0.81
0.8 to 1.2	696	1.00	0.10	0.10	1.01	0.34	0.34	1.07	0.57	0.53
1.2 to 1.6	133	1.34	0.10	0.07	1.51	0.41	0.27	1.37	0.69	0.50
1.6 to 2	73	1.73	0.10	0.06	1.87	0.40	0.22	1.91	0.67	0.35
>2	45	2.34	0.27	0.11	2.42	0.51	0.21	1.54	0.80	0.52
All Trips	1204	1.02	0.44	0.43	1.07	0.58	0.54	1.08	0.68	0.63

Table A.2: Headway Ratio at Outbound Time Points by Headway Ratio at Dudley

Headway Ratio at Harvard	Total Trips	Dudley			Hynes			Harvard		
		Mean	Std. Dev.	Coeff. of Var.	Mean	Std. Dev.	Coeff. of Var.	Mean	Std. Dev.	Coeff. of Var.
0 to 0.4	71	0.23	0.12	0.50	0.71	0.54	0.77	0.79	0.56	0.70
0.4 to 0.8	206	0.64	0.11	0.17	0.55	0.36	0.66	0.61	0.46	0.75
0.8 to 1.2	560	1.00	0.10	0.10	0.96	0.43	0.44	0.93	0.57	0.61
1.2 to 1.6	196	1.35	0.10	0.07	1.35	0.47	0.35	1.29	0.63	0.49
1.6 to 2	29	1.75	0.10	0.06	1.35	0.54	0.40	1.30	0.69	0.53
>2	28	3.00	0.88	0.29	1.31	1.16	0.88	1.82	0.65	0.36
All Trips	1090	1.02	0.48	0.47	0.96	0.54	0.56	0.96	0.62	0.65

Appendix B

Layover time distribution during experiment

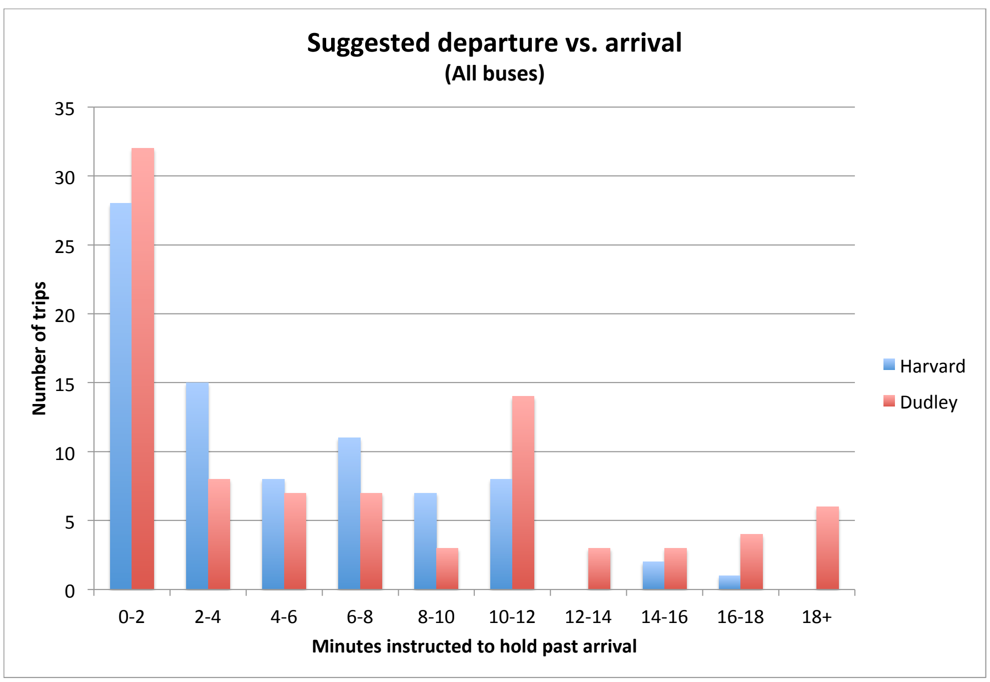

Figure B-1: Suggested Departure vs. Arrival Time at Each Terminal

Appendix C

Equivalent formulations of effective headway

Notation:

$\hat{\mu_h}$ = sample mean of observed headways

$\hat{\sigma_h}$ = sample standard deviation of observed headways

h_i = headway on trip i at the selected stop

N = the total number of trips observed

In this appendix, we demonstrate the equivalence of two formulae given in Section 6.1.2 for the effective headway metric.

$$\hat{\mu_h}\left(1+\left(\frac{\hat{\sigma_h}}{\hat{\mu_h}}\right)^2\right)=\frac{\sum\limits_i (h_i^2)}{\sum\limits_i h_i} \tag{C.1}$$

Beginning with the left-hand side, substitute the statistical definitions of $\hat{\mu_h}$ and $\hat{\sigma_h}$.

$$\hat{\mu_h}\left(1+\left(\frac{\hat{\sigma_h}}{\hat{\mu_h}}\right)^2\right)=\hat{\mu_h}\left(1+\frac{\hat{\sigma_h}^2}{\hat{\mu_h}^2}\right)=\frac{1}{N}\sum_i h_i\left(1+\frac{\frac{1}{N}\sum\limits_i h_i^2-\left(\frac{1}{N}\sum\limits_i h_i\right)^2}{\left(\frac{1}{N}\sum\limits_i h_i\right)^2}\right) \tag{C.2}$$

Then simplify as follows:

$$\frac{1}{N}\sum_i h_i \left(1 + \frac{\frac{1}{N}\sum\limits_i h_i^2 - \left(\frac{1}{N}\sum\limits_i h_i\right)^2}{\left(\frac{1}{N}\sum\limits_i h_i\right)^2}\right) = \frac{1}{N}\sum_i h_i + \frac{\sum\limits_i h_i^2 - \frac{1}{N}\left(\sum\limits_i h_i\right)^2}{\left(\sum\limits_i h_i\right)^2} = \frac{\sum\limits_i (h_i^2)}{\sum\limits_i h_i} \tag{C.3}$$

Appendix D

Weighted version of coefficient of variation

In this appendix, we determine the appropriate formula for the weighted coefficient of variation to use in our calculations. We follow a similar approach to that found in Appendix C, but this time in reverse, as we begin with the formula for effective headway, weighted by passenger arrival rate at each stop.

For convenience of notation, let $M = \sum_o \sum_i \lambda_p^o$. By a similar transformation to that used in Appendix C, we decompose the effective headway into the weighted mean $\hat{\mu_h} = \frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o$ and a component containing the squared coefficient of variation.

$$\begin{aligned} H_E = \frac{\sum_i \sum_o \lambda_p^o (h_i^o)^2}{\sum_i \sum_o \lambda_p^o h_i^o} &= \frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o \left(1 + \frac{\frac{1}{M} \sum_i \sum_o \lambda_p^o (h_i^o)^2 - \left(\frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o \right)^2}{\left(\frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o \right)^2} \right) \\ &= \hat{\mu_h} \left(1 + \frac{\frac{1}{M} \sum_i \sum_o \lambda_p^o (h_i^o)^2 - \left(\frac{1}{M} \sum_i \sum_o \lambda_p^o h_i^o \right)^2}{\hat{\mu_h}^2} \right) \end{aligned} \tag{D.1}$$

In order for Assumption 4 in Section 6.2.1 to hold, we must show that the sample mean headway $\hat{\mu}_h$ is approximately equal to the scheduled headway H_S. We begin by noting that, since the passenger arrival rate at each stop is assumed to be constant throughout the time period, λ_p^o is constant with respect to the bus trip i. Therefore, $M = \sum_o \sum_i \lambda_p^o = \sum_o \lambda_p^o \sum_i 1 =$

$N\sum_o \lambda_p^o$, where N is the number of trips. We can rewrite the weighted mean headway as follows:

$$\hat{\mu_h} = \frac{1}{N\sum_o \lambda_p^o} \sum_o \sum_i \lambda_p^o h_i^o = \hat{\mu_h} = \frac{1}{N\sum_o \lambda_p^o} \sum_o \left(\lambda_p^o \sum_i h_i^o \right) \tag{D.2}$$

Since the sum of the headways at a particular stop $\sum_i h_i^o$ is simply the time span from the first departure at that stop to the last departure, this value will be approximately constant across each origin stop o on a route. Treating $\sum_i h_i^o$ as a constant gives us the following:

$$\frac{1}{N\sum_o \lambda_p^o} \sum_o \left(\lambda_p^o \sum_i h_i^o \right) = \frac{1}{N\sum_o \lambda_p^o} \sum_i h_i^o \left(\sum_o \lambda_p^o \right) = \frac{1}{N} \sum_i h_i^o \tag{D.3}$$

Therefore, the weighted mean headway is approximately equal to the mean headway at any particular stop, which is approximately equal to the scheduled headway.

Bibliography

Abkowitz, M. D. & Lepofsky, M. (1990), 'Implementing headway-based reliability control on transit routes', *Journal of Transportation Engineering* **116**(1), 49–63.

Abkowitz, M., Slavin, R., Waksman, R., Englisher, L. & Wilson, N. (1978), Transit service reliability, Technical Report UMTA-MA-06-0049-78-l, U.S. Department of Transportation.

Bartholdi, J. J. & Eisenstein, D. D. (2012), 'A self-coördinating bus route to resist bus bunching', *Transportation Research Part B: Methodological* **46**(4), 481–491.

Carrel, A., Halvorsen, A. & Walker, J. L. (2013), 'Passengers' perception of and behavioral adaptation to unreliability in public transportation', *Transportation Research Record: Journal of the Transportation Research Board* **2351**(1), 153–162.

Cham, L. (2006), Understanding bus service reliability: A practical framework using avl/apc data, Master's thesis, Massachusetts Institute of Technology.

Chow, W., Block-Schachter, D. & Hickey, S. (2014), Impacts of real time passenger information signs in rail stations at the massachusetts bay transportation authority, *in* 'Transportation Research Board 93rd Annual Meeting', number 14-0194.

Chow, W. et al. (2014), Evaluating online surveys for public transit agencies using a prompted recall approach, Master's thesis, Massachusetts Institute of Technology.

Cosgrove, C. (2013), 'A cure for bus bunching'.
URL: *http://its.berkeley.edu/btl/2013/fall/busbunching*

Delgado, F., Munoz, J. C. & Giesen, R. (2012), 'How much can holding and/or limiting boarding improve transit performance?', *Transportation Research Part B: Methodological* **46**(9), 1202–1217.

Eberlein, X. J., Wilson, N. H., Barnhart, C. & Bernstein, D. (1998), 'The real-time deadheading problem in transit operations control', *Transportation Research Part B: Methodological* **32**(2), 77–100.

Eberlein, X. J., Wilson, N. H. & Bernstein, D. (1999), Modeling real-time control strategies in public transit operations, *in* 'Computer-aided transit scheduling', Springer, pp. 325–346.

Eberlein, X. J., Wilson, N. H. & Bernstein, D. (2001), 'The holding problem with real–time information available', *Transportation science* **35**(1), 1–18.

Ferris, B., Watkins, K. & Borning, A. (2010), Onebusaway: results from providing real-time arrival information for public transit, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, pp. 1807–1816.

Furth, P. G., Hemily, B., Muller, T. H. & Strathman, J. G. (2006), Using archived avl-apc data to improve transit performance and management, Technical Report Project H-28, Transportation Research Board.

Levinson, H. S. (1991), 'Supervision strategies for improved reliability of bus routes', *NCTRP Synthesis of Transit Practice* (15).

Lizana, P., Muñoz, J. C., Giesen, R. & Delgado, F. (2014), 'Bus control strategy application: Case study of santiago transit system', *Procedia Computer Science* **32**, 397–404.

Milkovits, M. N. (2008), Simulating service reliability of a high frequency bus route using automatically collected data, Master's thesis, Massachusetts Institute of Technology.

Moses, I. E. (2005), A transit route simulator for the evaluation of control strategies using automatically collected data, Master's thesis, Massachusetts Institute of Technology.

Pangilinan, C. (2006), Bus supervision deployment strategies for improved bus service reliability, Master's thesis, Massachusetts Institute of Technology.

Pangilinan, C., Wilson, N. & Moore, A. (2008), 'Bus supervision deployment strategies and use of real-time automatic vehicle location for improved bus service reliability', *Transportation Research Record: Journal of the Transportation Research Board* **2063**(1), 28–33.

Roxbury Dorchester Mattapan Transit Needs Study (2012), Technical report, Massachusetts Department of Transportation.

Sanchez-Martinez, G. (2012), Running time variability and resource allocation: A data-driven analysis of high-frequency bus operations, Master's thesis, Massachusetts Institute of Technology.

Sanchez-Martinez, G. (2014), 'Real-time holding control for high-frequency transit'.

Sanchez-Martinez, G. (2015), Real-time operations planning and control of high-frequency transit, PhD thesis, Massachusetts Institute of Technology.

Shen, S. & Wilson, N. H. (2001), An optimal integrated real-time disruption control model for rail transit systems, *in* 'Computer-aided scheduling of public transport', Springer, pp. 335–363.

Song, W. (1998), Real time dispatching control in transit systems, Master's thesis, Massachusetts Institute of Technology.

Strathman, J., Kimpel, T., Dueker, K., Gerhart, R., Turner, K., Griffin, D. & Callas, S. (2001), 'Bus transit operations control: Review and experience involving tri-met's automated bus dispatching system'.

Tempo Case Study: Bus synchronization in San Sebastian, Spain (2013), Technical report, VIA Analytics in partnership with Dbus Transit.

Transit Capacity and Quality of Service Manual (1999), Technical Report Project H-28, Transportation Research Board.

Tribone, D., Block-Schachter, D., Salvucci, F. P., Attanucci, J. & Wilson, N. H. (2014), 'Automated, data-driven performance regime for operations management, planning, and control', *Transportation Research Record: Journal of the Transportation Research Board* **2415**(1), 72–79.

Turnquist, M. A. (1981), 'Strategies for improving reliability of bus transit service', *Transportation Research Record* (818).

Turnquist, M. A. & Blume, S. W. (1980), 'Evaluating potential effectiveness of headway control strategies for transit systems', *Transportation Research Record: Journal of the Transportation Research Board* (746), 25–29.

Welding, P. (1957), 'The instability of a close-interval service', *OR* **8**(3), 133–142.

Wilson, N. H., Macchi, R., Fellows, R. & Deckoff, A. (1992), 'Improving service on the mbta green line through better operations control', *Transportation Research Record* (1361), 10–15.

Xuan, Y., Argote, J. & Daganzo, C. F. (2011), 'Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis', *Transportation Research Part B: Methodological* **45**(10), 1831–1845.